# COMPUTER SIMULATION OF SECONDARY STRUCTURE OF BIOLOGICAL AND SYNTHETIC MACROMOLECULES

A Thesis
Presented to
The Academic Faculty

by

Wei Zhang

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemical & Biomolecular Engineering

Georgia Institute of Technology
August 2009

# COMPUTER SIMULATION OF SECONDARY STRUCTURE OF BIOLOGICAL AND SYNTHETIC MACROMOLECULES

Approved by:


Professor Pete Ludovice, Advisor
School of Chemical & Biomolecular Engineering
*Georgia Institute of Technology*

Professor Athanassios Sambanis
School of Chemical & Biomolecular Engineering
*Georgia Institute of Technology*

Professor Rachel Chen
School of Chemical & Biomolecular Engineering
*Georgia Institute of Technology*

Professor Roger Wartell
School of Biology
*Georgia Institute of Technology*

Professor Steve Harvey
School of Biology
*Georgia Institute of Technology*


Date Approved: April 28th 2009

# ACKNOWLEDGEMENTS

I express my deepest gratitude to my advisor, Prof. Pete Ludovice, for his support, guidance, and wisdom during my here at Georgia Tech. I would like to acknowledge my committee members, Prof. Rachel Chen, Steve Harvey, Athanassios Sambanis, Roger Wartell. I would especially like to thank Prof. Loren Williams and Prof. Steve Harvey for their knowledge and guidance on the RNA project. I would also like to thank my group members and the group members of Prof. Steve Harvey and Loren Williams, Ayanna Bernard, Andrew Swann, Derrick Callandar, Dr. Anton Petrov, Tom Caulfield, Dr. Eli Hershkowitz, Chiaolong Hsiao, Srividya Mohan for camaraderie and collaboration.

I would like to thank my parents, Zhiqiang Zhang and Maomei Yang for their encouragement throughout my scholarly pursuits. Lastly I would like to thank my husband, Qing Tong who has been by my side through all of this. Without his unconditional love and unimaginable patience, my thesis work would not have been possible.

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS……..iii

LISTS OF TABLES……..…….....ix

LIST OF FIGURES……x

LIST OF SYMBOLS OR ABBREVIATIONS……...xxi

SUMMARY……...xxii

I INTRODUCTION……...1

1.1 Motivation……1

1.2 RNA……2

1.2.1 RNA Structure……2

1.2.2 RNA Tetraloops……..5

1.2.3 Structural Database……5

1.2.4 Standard Tetraloop and d2 Tetraloop……..7

1.3 ROMP Poly(N-undecyl-exo-norbornene-5,6-dicarboximide)……8

1.4 Molecular Modeling…….……..10

1.4.1 Molecular Dynamics……..……..10

1.4.2 Force Field……..……..11

1.4.3 Wide Angle X-ray Diffraction……..……12

1.5 Objectives and Hypotheses……13

1.5.1 Part I: RNA Tetraloops……13

1.5.2 Part II: ROMP PNBDC_UD……15

1.6 References……..16

II STANDARD TETRALOOP AND PERFECT D2 TETRALOOP……24

2.1 Introduction……24

2.1.1 Background……..24

2.1.1.1 RNA Deviations of Local Structure (DevLS)…….……...24

2.1.1.2 Standard Tetraloop and d2 Tetraloop……25
2.1.1.3 Nonexistence of Perfect d2 Tetraloop……...27
2.1.2 Motivation……29
2.1.3 Objectives……31
2.2 Methods……...…32
2.2.1 Modification of a Sample d2 Tetraloop with Stem Defects…….32
2.2.2 Simulation Methods……33
2.2.2.1 Molecular Dynamics (MD) Simulation……33
2.1.2.2 MD Simulation Protocol……35
2.3 Results and Discussions……36
2.3.1 Berendsen Temperature Coupling vs. Langevin Dynamics Temperature Control……36
2.3.2 Simulation Observations……..38
2.3.2.1 Simulation Results of Standard Tetraloop……38
2.3.2.2 Simulation Results of the Perfect d2 Tetraloop……..40
2.3.3 Quantitative Comparisons of Simulation Results between Standard Tetraloop and Perfect d2 Tetraloop……..43
2.3.3.1 RMSD Analysis……44
2.3.3.2 Hydrogen Bond and Base Pairing Tracking……...51
2.3.3.3 Torsion Angle Analysis……..62
2.4 Conclusions……69
2.5 References……..70
III STABILIZING EFFECTS OF STEM DEFECTS IN D2-TETRALOOPS……...73
3.1 Introduction……73
3.1.1 Background……..73
3.1.2 Motivation……74
3.1.3 Objectives……75
3.2 Methods……..76
3.2.1 Molecular Dynamics Simulation on 314-d2-tetraloop……76

3.2.2 Simulation Protocol……..77
3.3 Results and Discussions……..78
3.3.1 Simulation Results Comparison between the Perfect d2 Tetraloop and the d2 Tetraloop…….78
3.3.1.1 Simulation Observations……78
3.3.1.2 RMSD Analysis comparisons……..80
3.3.1.3 Hydrogen Bond and Base Pairing Tracking……..87
3.3.1.4 Torsion Angle Analysis of 314-d2-tetraloop……..93
3.3.2 Hydrogen Bonding of Insertion Residue 318C……96
3.3.3 Simulations of de316-314-d2-tetraloop…….100
3.3.3.1 Simulation observations……..…101
3.3.3.2 RMSD Analysis of de316-314-d2-tetraloop……103
3.3.3.3 Hydrogen Bonding Analysis of de316-314-d2-tetraloop……..105
3.3.4 Simulation Results of de316-de318-314-d2-tetraloop……..110
3.3.4.1 Simulation Observations……..111
3.3.4.2 RMSD Analysis of de316-de318-314-d2-tetraloop……..113
3.3.4.3 Hydrogen bonding analysis……..115
3.4 Conclusions……..……118
3.5 References……119
IV FOLDING PATHWAY OF TETRALOOP……..120
4.1 Introduction……..120
4.1.1 Background……120
4.1.1.1 The Hierarchical Folding of RNA Structure……120
4.1.1.2 Formation of RNA Hairpin Secondary Structure…….122
4.1.2 Motivation……..123
4.1.1.1 Functions of RNA Hairpins in Biology……..123
4.1.3 Objectives……...125
4.2 Methods……126
4.2.1 Secondary Structure of 2249-standard Tetraloop……..126

4.2.2 Thermal Unfolding Simulation……..127
4.2.3 Simulation Protocols……..128
4.3 Results and Discussions……...129
4.3.1 Unfolding Pathway 1……..130
4.3.1.1 RMSD Analysis of Unfolding Pathway 1……...130
4.3.1.2 Hydrogen bonding interactions…….134
4.3.1.3 High Temperature Unfolding Trajectory 1……..139
4.3.2 Folding Pathway 2……..141
4.3.2.1 RMSD Analysis of Unfolding Simulation 2……142
4.3.2.2 Hydrogen Bonding Interactions…….146
4.3.1.3 High Temperature Unfolding Trajectory 2……..151
4.4 Conclusions……...153
4.5 References……154
V ROMP POLY(N-UNDECYL-*EXO*-NORBORNENE-5,6-DICARBOXIMIDE)
…….158
5.1 Introduction……..158
5.1.1 Background……158
5.1.1.1 Polymer as support in immobilized catalysts……..158
5.1.1.2 Poly(norbornene)……160
5.1.1.3 ROMP poly(N-undecyl-exo-norbornene-5,6-dicarboximide)…….165
5.1.1.4 Wide angle X-ray diffraction (WAXD)……...166
5.1.2 Motivation……..169
5.1.3 Objectives……...170
5.2 Methods……171
5.2.1 Starting Conformation……171
5.2.2 Molecular Dynamics……...……...172
5.2.3 Wide Angle X-Ray Diffraction……...……....172
5.3 Results and Discussions……...173
5.3.1 Isolated Polymer Simulation……..173

5.3.1.1 *Cis*-poly(acetylene) and *Cis*-t-butyl poly(acetylene)…….173
5.3.1.2 ROMP PNBDC Simulations…….176
5.3.1.2 ROMP PNBDC_UD Simulations……..178
5.3.2 Bulk simulation……..182
5.3.2.1 ROMP PNBDC_UD1 at 300K……..182
5.3.2.2 ROMP PNBDC_UD1 at 900K……..183
5.3.2.1 ROMP PNBDC_UD 2 at 300K…….185
5.3.3 Wide angle X-ray diffraction……..187
5.3.3.1 WAXD of ROMP PNBDC_UD 1……..188
5.3.3.2 ROMP PNBDC_UD 2…….190
5.4 Conclusions……..192
5.5 References……193
VI CONCLUSIONS AND RECOMMENDATIONS……197
6.1 Conclusions……..197
6.2 Recommendations and Future work…….200
6.3 References……202

# LIST OF TABLES

2.1 805-standard tetraloop hydrogen bonding statistics for each base pair in the stem……...54

2.2 Cutd2-tetraloop hydrogen bonding statistics for each base pair in the stem…….58

2.3 Bin assignments and torsion ranges……..63

2.4 ASCII symbols, bin numbers and observation frequencies……..64

2.5 Tetraloop identified by the binning method……..65

2.6 Four backbone torsion angle for four residues in the loop of 805-standard tetraloop…….66

2.7 Four backbone torsion angle for three residues in the loop of cutd2-tetraloop….68

3.1 All residues, stem and loop RMSD statistics for 805-standard tetraloop, 314-d2-tetraloop and cutd2-tetraloop……83

3.2 All three stem base pairs RMSD statistics for 805-standard tetraloop, 314-d2-tetraloop and cutd2-tetraloop……86

3.3 314-d2-tetraloop hydrogen bonding statistics for each base pair in the stem…...90

3.4 Four backbone torsion angle for three residues in the loop of 314-d2-tetraloop..94

# LIST OF FIGURES

1.1 RNA nucleotide structure and four bases, adenine, guanine, cytosine and uracil.....3

1.2 Structure and residue positions nomenclature of both standard tetraloop (left) and d2 tetraloop (right)……..…..……...7

1.3 Structure of ROMP PNBDC_UD……...10

1.4 WAXD spectra for PNB. The dark solid line represents the experimental WAXD pattern for PNB, the dashed line represents a four chain model of PNB, the solid line represents a six chain model of PNB, and the dotted line represents the WAXD pattern for a poly(ethylene)-PNB 50:50 random copolymer……...13

2.1 RNA DevLS. A) A generic RNA motif is represented schematically by four circles, which symbolize four residues. B) In a motif with 3-2 switch, the positions of two bases, of residues 3 and 2 in the figure, are interchanged. The backbone linkage is maintained. C) In a deleted motif, a residue is omitted (dashed line). D) In an inserted motif, a residue is added. E) In a strand clipped motif one or more residues is contributed from a remote region of the primary sequence. An insertion, if extensive enough can be equivalent to a strand clip……25

2.2 Residue position nomenclature and secondary structure of a standard tetraloop……26

2.3 Residue position nomenclature and secondary structure of a d2 tetraloop……..…...27

2.4 Secondary structure of the HM 23S rRNA (1JJ2). Tetraloop locations and type are indicated by color. Superscripted c's indicate strand clipped

tetraloops…….……...28

2.5 A rough sketch of the secondary structure of 314-d2-tetraloop (left) and cutd2-tetraloop (right)….……...33

2.6 A rough sketch of the secondary structure of 805-standard tetraloop……...……...34

2.7 The structure of 805-standard tetraloop with hydrogen bond highlighted in dotted line at the end of 10ns simulation……..……..39

2.8 The snapshots taken from the MD simulation of 805-standard tetraloop, the left figure is the starting point, and the right figure is the end structure of 10ns of simulation……..40

2.9 The snapshots from the MD simulation of cutd2-tetraloop, the left figure is the starting point, and the right figure is the structure at the end of the simulation…41

2.10 The end structures of 805 standard tetraloop (left) and cutd2-tetraloop (right) from the simulation……....43

2.11 All residues RMSD calculation for 805-standard tetraloop (purple solid line) and cutd2-tetraloop (blue dotted line)……45

2.12 Loop region RMSD calculations for 805-standard tetraloop (purple solid line) and cutd2-tetraloop (blue dotted line)……46

2.13 Stem region RMSD calculations for 805-standard tetraloop (purple solid line) and cutd2-tetraloop (blue dotted line)……47

2.14 The terminus base pair RMSDs calculated over 10 ns of simulation for both the standard tetraloop (805-standard tetraloop in solid purple line) and the perfect d2 tetraloop (cutd2-tetraloop in dotted blue line)……...48

2.15 The central base pair RMSDs calculated over 10 ns of simulation for both the standard tetraloop (805 standard tetraloop in solid purple line) and the perfect d2 tetraloop (cutd2-tetraloop in dotted blue line)……...49

2.16 The closing base pair RMSDs calculated over 10 ns of simulation for both standard tetraloop (805 standard tetraloop in solid purple line) and the perfect

d2 tetraloop (cutd2-tetraloop in dotted blue line)……...49

2.17 The terminus base pair of 805 standard tetraloop. Green line is the atom distance between H1 on 802 Guanine and N3 on 811 Cytosine. Red line is the atom distance between O6 on 802 Guanine and H41 on 811 Cytosine. Blue line is the atom distance between H21 on 802 Guanine and O2 on 811 Cytosine……..52

2.18 The central base pair of 805 standard tetraloop. Green line is the atom distance between H41 on 803 Cytosine and O6 on 810 Guanine. Red line is the atom distance between O2 on 803 Cytosine and H21 on 810 Guanine. Blue line is the atom distance between N3 on 803 Cytosine and H1 on 810 Guanine……..53

2.19 The closing base pair of 805 standard tetraloop. Green line is the atom distance between H41 on 804 Cytosine and O6 on 809 Guanine. Red line is the atom distance between O2 on 804 Cytosine and H21 on 809 Guanine. Blue line is the atom distance between N3 on 804 Cytosine and H1 on 809 Guanine……..53

2.20 The terminus base pair of cutd2-tetraloop. Green line is the atom distance between O2 on 311 Cytosine and H21 on 319 Guanine. Red line is the atom distance between H41 on 311 Cytosine and O6 on 319 Guanine. Blue line is the atom distance between N3 on 311 Cytosine and H1 on 319 Guanine….56

2.21 The central base pair of cutd2-tetraloop. Red line is the atom distance between H3 on 312 Uracil and N1 on 318 Adenine. Blue line is the atom distance between O4 on 312 Uracil and H61 on 318 Adenine……...56

2.22 The closing base pair of 314-d2-tetraloop. Red line is the atom distance between H3 on 313 Uracil and N7 on 317 Adenine. Blue line is the atom distance between O4 on 313 Uracil and H61 on 317 Adenine……...57

2.23 Intra-loop hydrogen bonding interaction (j-1 and j+1) comparison between 805-standard tetraloop (blue) and cutd2-tetraloop (red). The distance between H2' on Guanine of residue 805 at position j-1 and N7 on Adenine of residue

807 at position j+1 in blue. The distance between H2' on Guanine of residue 314 at position j-1 and N7 on Adenine of residue 316 at position j+1 in red……...60

2.24 Intra-loop hydrogen bonding interaction (j-1 and j+2) comparison between 805-standard tetraloop (blue) and cutd2-tetraloop (red). The distance between H21 on Guanine of residue 805 at position j-1 and O2P on Adenine of residue 808 at position j+2 in blue, the distance between H1 on Guanine of residue 314 at position j-1 and O2P on Adenine of residue 317 at position j+2 in red……...61

2.25 Ribonucleotide torsion angles used to specified conformation……..62

3.1 Secondary structure of 314-d2-tetraloop……77

3.2 The snapshots of the start structure (left) and the end structure (right) from 10ns MD simulation of 314-d2-tetraloop……...78

3.3 The ending structures of the 314-d2-tetraloop (left) and the cutd2-tetraloop (right) from the simulation……80

3.4 All residues RMSD calculation for 314-d2-tetraloop (pink solid line) and cutd2-tetraloop (blue dotted line)……...81

3.5 Loop RMSD calculation for 314-d2-tetraloop (pink solid line) and cutd2-tetraloop (blue dotted line)……...82

3.6 Stem RMSD calculation for 314-d2-tetraloop (pink solid line) and cutd2-tetraloop (blue dotted line)……...82

3.7 The terminus base pair RMSDs calculated over 10 ns of simulation for both the d2 tetraloop (314-d2tetraloop in solid pink line) and the perfect d2 tetraloop (cutd2-tetraloop in dotted blue line)……84

3.8 The central base pair RMSDs calculated over 10 ns of simulation for both d2 tetraloop with bulge (314-d2tetraloop in solid pink line) and perfect d2 tetraloop (cutd2-tetraloop in dotted blue line)……85

3.9 The closing base pair RMSDs calculated over 10 ns of simulation for both d2 tetraloop with bulge (314-d2tetraloop in solid pink line) and perfect d2

tetraloop (cutd2-tetraloop in dotted blue line)……85

3.10 The terminus base pair of 314-d2-tetraloop. Green line is the atom distance between O2 on 311 Cytosine and H21 on 320 Guanine. Red line is the atom distance between H41 on 311 Cytosine and O6 on 320 Guanine. Blue line is the atom distance between N3 on 311 Cytosine and H1 on 320 Guanine….88

3.11 The central base pair of 314-d2-tetraloop. Red line is the atom distance between H3 on 312 Uracil and N1 on 319 Adenine. Blue line is the atom distance between O4 on 312 Uracil and H61 on 319 Adenine……..88

3.12 The closing base pair of 314-d2-tetraloop. Red line is the atom distance between H3 on 313 Uracil and N7 on 317 Adenine. Blue line is the atom distance between O4 on 313 Uracil and H61 on 317 Adenine……..89

3.13 Intra-loop hydrogen bonding interaction (j-1 and j+1) comparison between 314-d2-tetraloop (Solid red line) and cutd2-tetraloop (dotted blue line). The distance between H2' on Guanine of residue 314 at position j-1 and N7 on Adenine of residue 316 at position j+1 for both 314-d2-tetraloop and cutd2-tetraloop……...91

3.14 Intra-loop hydrogen bonding interaction (j-1 and j+2) comparison between 314-d2-tetraloop (red solid line) and cutd2-tetraloop (blue dotted line). The distance between H1 on Guanine of residue 314 at position j-1 and O2P on Adenine of residue 317 at position j+2 for both 314-d2-tetraloop and cutd2-tetraloop……..92

3.15 Snapshot of 314 d2-tetraloop from the simulation with hydrogen bond highlighted in dotted line……97

3.16 Hydrogen bonding interactions between residue H2' of 316A (j+1) and cytosine of 318C, atom distance between 316A(H2') and 318C(C2) in blue, 316A(H2') and 318C(O2) in red, and 316A(H2') and 318C(N3) in purple..98

3.17 Hydrogen bond interaction between hydroxyl group of insertion residue 318C and the base of residue 317A in the stem, atom distance between 318C(H2') and 317A( N3)…….99

3.18 Modification of 314-d2-tetraloop by removing the hydroxyl group (2') from the sugar ring of residue 316A. The structure on the left represents the original 314-d2-tetraloop with two hydrogen bonding of the insertion residue 318C (dash line). The structure on the right represents the deoxidized 314-d2-tetraloop, labeled de316-314-d2-tetraloop with only the hydrogen bonding between insertion residue 318C and the stem remaining……...101

3.19 Snapshots of de316-314-d2-tetraloop at 1.5ns of the simulation (left), which illustrated that the insertion was attempting to form hydrogen bonding with the loop, and at the end of the simulation (right)…….102

3.20 All atoms RMSD calculations of de316-314-d2-tetraloop over 10ns of simulation time…….103

3.21 Loop (blue) and stem (red) region RMSD calculations of de316-314-d2-tetraloop over 10ns of simulation time…….104

3.22 Distance calculated between atoms in the loop and the insertion residue that were observed to form hydrogen bonding in the simulation of de316-314-d2-tetraloop. Blue represents the atom distance between N3 on residue 316A and H41 on residue 318C, red represents the atom distance between H21' on residue 316A and N4 on residue 318C, and green represents the atom distance between N3 on residue 316A and H42 on residue 318C.106

3.23 The intra-loop hydrogen bonding interaction of residue j-1 and residue j+1 (in red) was compared to the hydrogen bonding interaction between residue 316A in the loop and the insertion residue 318C (in blue) and the hydrogen bonding interaction between N3 on residue 316A and H42 on residue 318C…….107

3.24 The hydrogen bonding interaction between N3 on residue 317A and H2' on residue 318C calculated over 10ns of the simulation……...109

3.25 Modification of de316-314-d2-tetraloop by removing hydroxyl group (2') from sugar ring of residue 318C. The structure on the left represents the original de316-314-d2-tetraloop with the hydrogen bonding interaction

between the insertion residue 318C and the stem. The structure on the right represents the deoxidized de316-314-d2-tetraloop, labeled de316-de318-314-d2-tetraloop with both hydrogen bonding of insertion residue 318C removed……..111

3.26 Snapshots of de316-de318-314-d2-tetraloop from 10ns of MD simulation. The left figure is the structure at the beginning of the simulation, and the right figure is the structure at the end of the simulation……...112

3.27 All atoms RMSD calculations of de316-de318-314-d2-tetraloop over 10ns of simulation time…….114

3.28 Loop (red) and stem (blue) region RMSD calculations of de316-de318-314-d2-tetraloop over 10ns of simulation time……..115

3.29 Distance calculated for atoms in the loop – insertion hydrogen bonding previously observed in the simulation of de316-314-d2-tetraloop, distance between N3 on residue 316A and H41 on residue 318C in blue, H21' on residue 316A and N4 on residue 318C in red, and N3 on residue 316A and H42 on residue 318C in green……..116

3.30 Intra-loop hydrogen bonding calculation between residue j-1 and residue j+1…….117

4.1 Stem loop sketches of hairpin structure……...122

4.2 Secondary structure of 2249-standard tetraloop……..126

4.3 All atoms RMSD calculations for high temperature simulation 1……...131

4.4 RMSD calculations of the loop region (red) and the stem (green) over 2ns of simulation time for high temperature unfolding simulation 1……..132

4.5 RMSD calculations for three base pair in the stem over 2ns of simulation time for unfolding simulation 1. The terminus base pair (2246U-2255A) in red, the central base pair (2247C-2254G) in blue, and the closing base pair (2248G-2253C) pair in green……...133

4.6 Hydrogen bonding interactions between the terminus base pair in the stem of unfolding simulation 1. Distance of O4 on 2246U and H61 on 2255A in pink,

distance of H3 on 2246U and N1 on 2255A in blue……134

4.7 Hydrogen bonding interactions between the central base pair in the stem. Distance of N3 on 2247C and H1 on 2254G in purple, distance of H41 on 2247C and O6 on 2254G in pink, and distance of O2 on 2247C and H2 on 2254G in blue……...135

4.8 Hydrogen bonding interactions between the closing base pair in the stem. Distance of H1 on 2248G and N3 on 2253C in purple, distance of O6 on 2248G and H41 on 2253C in pink, and distance of H21 on 2248G and O2 on 2253C in blue……...135

4.9 Intra-loop hydrogen bonding interactions between 2249G (j-1) and 2251G (j+1) from unfolding simulation 1……137

4.10 Intra-loop hydrogen bonding interactions between 2249G (j-1) and 2252A (j+2) from unfolding simulation 1, atom distance between H1 of 2249G and O2P of 2252A in pink and atom distance between H21 of 2249G and O2P of 2252A in blue……...138

4.11 Two nanosecond high temperature (510K) simulation trajectory 1…….140

4.12 All atoms RMSD calculations from high temperature simulation 2……142

4.13 RMSD calculations of the loop region (red) and the stem (blue) over 2ns of simulation time for high temperature unfolding simulation 2….....143

4.14 RMSD calculations of residue j-1, j, and j+1 in the loop region from unfolding simulation 2…….144

4.15 RMSD calculations for all three base pairs in the stem for unfolding simulation 2, closing base pair (2248G – 2253C) in green, central base pair (2247C – 2254G) in blue, and terminus base pair (2246U – 2255A) in red…….145

4.16 Hydrogen bonding interactions between the terminus base pair in the stem during unfolding simulation 2. Distance of O4 on 2246U and H61 on 2255A in purple, distance of H3 on 2246U and N1 on 2255A in blue……147

4.17 Hydrogen bonding interactions between the central base pair in the stem.

Distance of N3 on 2247C and H1 on 2254G in purple, distance of H41 on 2247C and O6 on 2254G in pink, and distance of O2 on 2247C and H2 on 2254G in blue……...147

4.18 Hydrogen bonding interactions between the closing base pair in the stem. Distance of H1 on 2248G and N3 on 2253C in purple, distance of O6 on 2248G and H41 on 2253C in pink, and distance of H21 on 2248G and O2 on 2253C in blue……...148

4.19 Intra-loop hydrogen bonding interactions between 2249G (j-1) and 2251G (j+1) for unfolding simulation 2……..149

4.20 Intra-loop hydrogen bonding interactions between 2249G (j-1) and 2252A (j+2), atom distance between H1 of 2249G and O2P of 2252A in green and atom distance between H21 of 2249G and O2P of 2252A in blue……..150

4.21 Snapshots from unfolding simulation 2. The folded state (starting point) is on the left, the intermediate state (at 500ps) is at the center, the unfolded state (at 1ns) is on the right. The phosphate-ribose backbones of the stem are displayed as blue sticks for clarity. The phosphate-ribose backbones of the loop are colored according to their residue ID, 2249G (j-1) in pink, 2250G (j) in blue, 2251G (j+1) in purple, and 2252A (j+2) in green……...151

5.1 A rough sketch of an immobilized homogeneous catalyst……..159

5.2 The three polymerization pathways of norbornene……..161

5.3 The different possible configurations of ROMP PNB. ROMP PNB can have (a) racemic or meso diads, can be (b) cis or trans across the double bond or (c) the cyclopentane ring, and can have (d) an attached sidechain on the cyclopentane ring…….164

5.4 Structure of ROMP poly(N-undecyl-exo-norbornene-5,6-dicarboximide)..166

5.5 WAXD spectra for PNB. (a) The dark solid line represents the experimental WAXD pattern for PNB, the dashed line represents a four chain model of PNB, the solid line represents a six chain model of PNB, and the dotted line represents the WAXD pattern for a poly(ethylene)-PNB 50:50 random

copolymer [16]. (b) WAXD spectra for MW from $8.5X10^3$ (top solid line) to $4.74X10^5$ (lowest dotted line) g $mol^{-1}$. ……167

5.6 Comparison between simulated and experimental WAXD pattern (shifted for comparison) for vinylic PNB with methyl, butyl, and hexyl side groups on the norbornene ring. The simulated and experimental WAXD pattern for the methyl, butyl, and hexyl PNB are represented by the solid line, short-long dashed, and dotted line, respectively……169

5.7 Simulation snapshots of *cis*-poly(acetylene). *(a)* helical starting structure, *(b)* structure after minimization, and *(c)* structure after 100ps of MD simulation…….174

5.8 Simulation snapshots of *cis*-t-butyl poly(acetylene). *(a)* helical starting structure, *(b)* structure after minimization, and *(c)* structure after 500ps of MD simulation……..175

5.9 End structure of ROMP PNBDC without the undecyl side group on the dicarboximide ring after 500ps MD…….177

5.10 Simulation snapshots of ROMP PNBDC_UD1. *(a)* helical starting structure, *(b)* structure after minimization (top structure being all atoms and bottom structure being backbone only), and *(c)* structure after 500ps of MD simulation (left structure being all atoms and right structure being backbone only)…….180

5.11 Simulation snapshots of ROMP PNBDC_UD2. *(a)* helical starting structure, *(b)* structure after minimization (top structure being all atoms and bottom structure being backbone only), and *(c)* structure after 500ps of MD simulation (left structure being all atoms and right structure being backbone only)…….181

5.12 Bulk simulation of ROMP PNBDC_UD1. The snapshot of backbone atoms after 500ps MD at 300K, density 0.9 g $cm^{-3}$……182

5.13 High temperature bulk simulation of ROMP PNBDC_UD1, backbone snapshot after 500ps of MD at 900K, density 0.9 g $cm^{-3}$…….183

5.14 Bulk simulation of ROMP PNBDC_UD2. The snapshot of backbone after 500ps MD at 300K, density 0.9 g $cm^{-3}$……186

5.15 WAXD pattern generated by Weck group experimentally, low angle peak at 2.5° , and high angle peak at 12.6° ……...188

5.16 WAXD pattern of ROMP PNBDC_UD1 after bulk simulation at 300K. Only a sharp peak at low angle (around 3° – 4°). No peak was observed at the high angle…….189

5.17 WAXD pattern of ROMP PNBDC_UD1 after bulk simulation at 900K. Only a sharp peak at low angle (around 3° – 4°). No peak was observed at the high angle…….189

5.18 Averaged WAXD pattern from 8 repeat of bulk simulations of ROMP PNBDC_UD2 (blue) and WAXD pattern from experiment (red). The low angle peak is at 2-3°, and the high angle peak is at 11.5° for simulated WAXD pattern, while the low angle peak is at 2.5°, and the high angle peak is at 12.6° for experimental WAXD pattern…….191

# LIST OF SYMBOLS AND ABBREVIATIONS

| | |
|---|---|
| DevLS | Deviation of Local Structure |
| DNA | Deoxyribonucleic Acid |
| HM | *Haloarcula marismortui* |
| LPTL | Lonepair Triloop |
| LSU | Large Subunit of ribosome |
| MD | Molecular Dynamics |
| miRNA | MicroRNA |
| mRNA | Messenger RNA |
| NPA | Nosé–Poincaré–Andersen |
| NPT | Isothermal-isobaric ensemble. Number of particles, temperature, and pressure are all held constant. |
| NVT | Canonical ensemble. Number of particles, system volume, and temperature are all held constant. |
| PDB | Protein Data Bank |
| PME | Particle Mesh Eward |
| PNB | Poly(norbornene) |
| PNBDC | poly(norbornene-5,6-dicarboximide) |
| PNBDC_UD | Poly(N-undecyl-exo-norbornene-5,6-dicarboximide) |
| RMSD | Root Mean Square Deviation |
| RNA | Ribonucleic Acid |
| ROMP | Ring-Opening Metathesis Polymerization |
| rRNA | Ribosomal RNA |
| tRNA | Transfer RNA |
| WAXD | Wide Angle X-ray Diffraction |

# SUMMARY

RNA tetraloop is the smallest, simplest and the most frequent motif which is involved in numerous important biological functions. A local deviation from the RNA standard tetraloop, d2 tetraloop, has been identified with high abundance in 5S, 16S and 23S rRNAs. The presence of d2 tetraloops in highly conserved regions of 16S and 23S rRNAs suggests their functional importance.

With one less residue in the loop, d2 tetraloops are considered more energetically constrained and less stable than standard tetraloops. The deletion at position j+2 in the loop is always correlated with an adjacent stem distortion. MD simulations of 314-d2-tetraloop (a sample structure of d2 tetraloops) and cutd2-tetraloop (an artificially built perfect d2 tetraloop with no stem defects) have shown that stem defects are the stabilizing factor of d2 tetraloops. Simulations have also revealed that the insertion residue 318C (an example of stem defect) is stabilizing 314-d2-tetraloop by forming hydrogen bonding interactions with both the loop and the stem. When these two hydrogen bonding interactions are eliminated, the structure remained relatively stable compared to cutd2-tetraloop where the insertion residue was completely removed from the stem. This suggests the insertion residue is also stabilizing 314-d2-tetraloop by providing some conformational relaxation in the stem.

Investigation of RNA standard tetraloop high temperature unfolding has revealed that the d2 tetraloop is possibly a kinetically trapped intermediate state during the folding of the standard tetraloop. High temperature unfolding simulations of standard tetraloop have shown a three-state folding behavior: a folded state, an intermediate state and an unfolded state. The folding of standard tetraloop starts with the formation of the loop. The closing base pair forms first, followed by the loop and the stem which form critical

interactions such as base pairing and stacking that make a tetraloop.

ROMP PNB has been investigated as supports for immobilized homogeneous catalysts to achieve high reactivity and high selectivity and easy separation. Polymers with intermediate conformational order can increase the accessibility of tethered homogeneous catalysts. Simulations of ROMP PNBDC_UD have shown the importance of bulky side groups in enabling the polymer to adopt a helical conformation. Such helical conformations have been associated with intermediate structural order in similar polymers such as PNB made by non-ROMP mechanisms. This intermediate order manifests itself as a split in the amorphous halo of WAXD pattern. Bulk simulations of ROMP PNBDC_UD generated WAXD patterns that are close to the experimentally generated WAXD patterns where there are two split peaks: lower angle peak representing intermolecular interaction and higher angle peak representing intramolecular interaction. This indicates that the imide and side group are required to induce helices.

# CHAPTER I

# INTRODUCTION

## *1.1 Motivation*

Computer simulations have been widely used in the fields of computational biology and material science for studying molecular systems ranging from biological molecules to synthetic polymers. Computer simulations can be considered as virtual experiments that are between theoretical models and experimental data collected in the laboratory. Computer simulations will not replace experimental work, but rather supplement the experimental work by explaining results of the experiments and providing information that cannot be obtained via experiment. Molecular Dynamics simulation, as a form of computer simulation, has become a powerful predictive tool to investigate the structures, dynamics and thermodynamics of biological and synthetic polymer systems.

This work applies Molecular Dynamics (MD) as the principle tool to study the structures of both a biological molecule and a synthetic polymer: Ribonucleic Acid (RNA) and ROMP Poly(N-undecyl-*exo*-norbornene-5,6-dicarboximide) (ROMP PNBDC_UD). RNAs are structurally very flexible and can adopt numerous conformations that are not well understood. This work focused on the study of RNA tetraloops, one of the smallest, simplest and most frequent motifs. MD simulations performed on this small RNA structure can provide great insight on its thermodynamics stability and its folding pathway.

Traditionally synthetic polymers form either a regular crystal or an amorphous random coil. However, there is a class of polymer that forms an intermediate level of order: a helix-kink conformation. A good example of such a polymer is vinylic poly(norbornene). Its ROMP cousin, ROMP poly(norbornene) forms a random coil due to the high flexibility of the cyclopentane ring on its backbone. Simulations performed on ROMP PNBDC_UD, a ROMP PNB with bulky side groups added to its backbone to increase rigidity can reveal the role of bulky side groups in making the polymer adopting an intermediate order.

## *1.2 RNA*

### 1.2.1 RNA Structure

Ribonucleic Acid (RNA) consists of a long chain of nucleotide units. Each nucleotide consists of a base, a ribose sugar and a phosphate. The ribose sugar has five carbons numbered 1' through 5'. A base is attached to the 1' position, generally adenine (A), cytosine (C), guanine (G) or uracil (U). A phosphate group is attached to the 3' position of one ribose and the 5' position of the next (Figure 1.1).

RNA has long been underestimated in its capacities, originally considered an uninteresting contaminant of the more important proteins. Since 1950s, the involvement of ribosomal RNA (rRNA) transfer RNA (tRNA) and messenger RNA (mRNA) in many important biological processes like translation, transcription, regulation and protein synthesis was gradually discovered [1, 2]. As a further surprise, in 1975 Thomas Cech [3] and colleagues showed that RNA could also be catalytically active in addition to carrying genetic information. In view of these additional capabilities, RNA is now generally considered to have been critical for the evolution of life before the genetic code had evolved: RNA could thus have performed dual functions in ancestral cells for which DNA and proteins are now used [4].

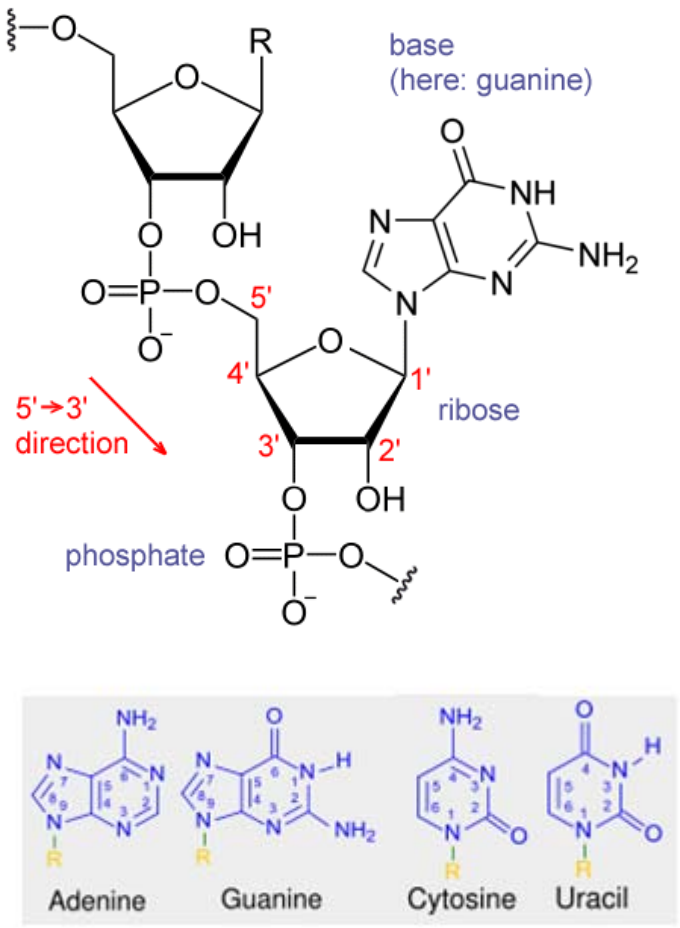



Figure 1.1: RNA nucleotide structure and four bases, adenine, guanine, cytosine and uracil [5, 6].

This functional diversity is also reflected in the structural diversity of RNA. Its structural polymorphism arises from the large number of degrees of freedom of its backbone, which has six backbone torsion angles and one angle describing the rotation of the base relative to the sugar [7]. The structural polymorphism of RNA gives rise to a variety of RNA structural motifs.

It was believed that prediction and design of three-dimensional structures of large RNAs are best approached using small structural motifs, with modular and hierarchical characteristics [8, 9]. The term RNA motif has different definitions. Moore

[9] defined an RNA motif as a discrete sequence or combination of base juxtapositions found in naturally occurring RNAs in unexpectedly high abundance. All of these discrete, hyperabundant sequences and base juxtapositions have distinctive conformations that are independent of the context in which they are embedded. Westhof [10] considered RNA motifs to be ordered stacked arrays of base pairs with distinctive backbone geometries with ordered arrays of isosteric non-Watson-Crick base pairs. RNA motifs are also defined to have known binding or chemical activity [11].

### 1.2.2 RNA Tetraloops

Among these motifs, Tetraloops are the simplest, smallest and most frequent motif and are the focus of the biological macromolecules part of this study (Chapter II, III and IV). Tetraloops are terminal loops, which were first observed in early phylogenetic comparisons of RNAs as characteristic four residue sequences [12–14]. A terminal loop is any sequence where an RNA folds back on itself so that a stem can form [9]. Tetraloops have been broadly grouped by sequence into three classes, which are GNRA [15-19], UNCG [14, 20-24], and CUUG [25, 26], (where N can be any nucleotide and R is either G or A), and have extraordinary thermodynamic stabilities [14, 27, 28].

Tetraloops have a lot of important biological functions and are thought to initiate folding of complex RNA molecules [14], to stabilize helical stems [14, 29], and to provide recognition elements for tertiary interactions and protein binding [15, 30, 31]. Therefore, understanding the stability and the folding mechanism of tetraloops would expand the overall understanding of numerous important biological functions of this particular RNA structure.

### 1.2.3 Structural Database

The structural databases, e.g. Nucleic Acid Database [32], Protein Data Bank (PDB) [33], which contain 3D structures of DNA, RNA and protein molecules, have grown explosively in the last decade. To date, more than 50 thousand structures have been deposited in the Protein Data Bank, with about four thousand nucleic acid structures in the Nucleic Acid Database.

The crystal structures of very large RNA molecules are now available at high resolution. At 2.4 Å resolution, there is a large subunit (LSU) of the ribosomal RNA from *archaeon H. marismortui* [34, 35] (PDB entry 1JJ2), which is a halophile from the Dead Sea. At 2.8 Å resolution, there is a ribosomal RNA assembly from *eubacterium T. thermophilus* [36] (PDB entry 2J00, 2J01), isolated from a thermal vent. At 2.3 Å resolution, there is a 160 nucleotide P4 – P6 domain of the self-slicing *Tetrahymena thermophila* group I intron [37] (PDB entry 1HR2). These are large structures rich in conformation, folding, tertiary base – base interactions, secondary structures and kinetically trapped intermediates and are therefore valuable database for studying RNA structures.

The HM 23S rRNA (1JJ2) is the test 'database' for this study and is generally referred to as the database in the three chapters on RNA. At 2.4 Å resolution, the atomic positions of the vast majority of the 23S rRNA of HM LSU are well-characterized and are more acutely determined than any other large RNA complex. The HM 23S rRNA, with over 2500 residues, constitutes a large database for studying RNA conformation and interactions.

### 1.2.4 Standard Tetraloop and d2 Tetraloop

Williams and Coworkers [38] show that RNA tolerates variation in conformation, topology, and molecular interactions. The variation is described by four distinct modes, which are insertions, deletions, strand clips and 3-2 switches, collectively called DevLS (Figure 2.1). In their study, Williams and coworkers focused on the tetraloop motif aiming at providing broad implications for RNA structure in general.

In collaboration with Williams group, we focus our study on the standard tetraloop and the d2 tetraloop, which belongs to the deletion family of DevLS. Figure 1.2 shows the structure and nomenclature of residue positions of both standard tetraloop and d2 tetraloop.

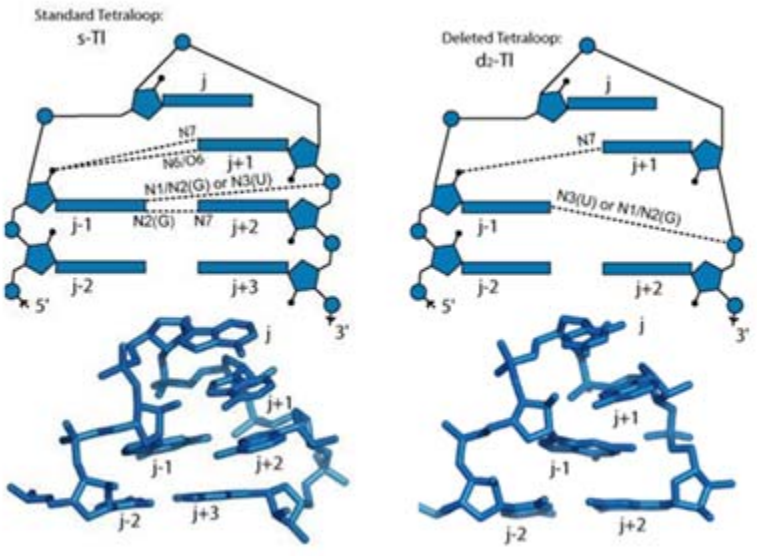



Figure 1.2: Structure and residue positions nomenclature of both standard tetraloop (left) and d2 tetraloop (right) [38].

In a d2 tetraloop, residue j+2 of a standard tetraloop is absent. Residue j+3 of a standard tetraloop becomes j+2 of a d2 tetraloop. The deletion of residue j+2 in the loop area does not appreciably change the positions of the remaining backbone atoms of these tetraloops. The d2 tetraloops are more strained in conformation than the standard tetraloops due to the fact that there are only three residues in the loop region. Out of 40 identified tetraloops from the 23S rRNA of HM LSU database, 21 are standard tetraloops and 10 are d2 tetraloops. The reason for the high abundance of d2 tetraloop in the database is unknown.

Williams and coworkers found that the deletion at position j+2 is always correlated with adjacent helical distortions such as insertions at position 3, clipping at position 2, or base pairing disruption in the stem. In other words, a perfect d2 tetraloop with no stem distortion has not been observed in the database. Our hypothesis is that the stem distortion plays the role of relaxing the stem to stabilize the d2 tetraloop.

## *1.3 ROMP Poly(N-undecyl-exo-norbornene-5,6-dicarboximide)*

Polymers, as homogeneous catalysts supports, have previously been used in the application of immobilized homogeneous catalysis and provide many advantages as supports for such systems [39]. There are a number of types of polymeric chains that can be used as catalysts supports. One such type includes linear chains with functionalized end groups. Linear chains with a catalyst either on every monomer unit or some other configuration that offers higher loading are also available for use. If a catalyst is tethered to a polymer support, a random coil polymer has the potential to block the catalyst site to the reactant. A random coil polymer has no underlying

structure so it is possible that the catalyst site could be on the interior of the polymer molecule. Therefore polymers with helical conformations with an expanded overall conformation are desirable in immobilized catalysis systems.

Modeling of poly(norbornene) created via ring-opening metathesis polymerization (ROMP PNB) by Ludovice group showed that all of the conformations of ROMP PNB form a random coil. It was determined that the presence of the random coil conformations was due to the fact that the cyclopentane ring on the backbone was not rigid enough to be considered as non-rotatable [40]. However, it is possible to functionalize the cyclopentane ring in such a manner as to remove the flexibility of the backbone cyclopentane ring.

Recent experimental studies have shown results that may suggest that the polymerization of a repeat unit with the addition of a second ring fused to the cyclopentane ring could produce a polymer that has a helix-kink conformation [41–45]. ROMP poly(N-undecyl-*exo*-norbornene-5,6-dicarboximide) (PNBDC_UD) is an example of PNB with such a bulky side group (Figure 1.3). In ROMP PNBDC_UD, a second ring is fused to the backbone cyclopentane ring of ROMP PNB to make the cyclopentane ring more rigid and therefore behave as a non-rotatable bond, thus producing a helical conformation.

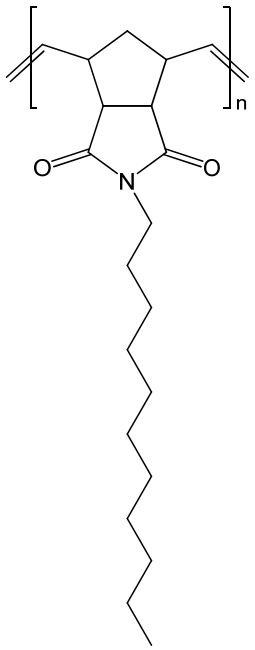



Figure 1.3: Structure of ROMP poly(N-undecyl-*exo*-norbornene-5,6-dicarboximide).

## *1.4 Molecular Modeling*

Molecular modeling can be a powerful tool and is now routinely used to investigate the structure, dynamics and thermodynamics of biological and polymeric systems. It will not replace experimental work, but rather supplement the experimental work by explaining results of the experiments and providing information that cannot be obtained via experiment.

### 1.4.1 Molecular Dynamics

One of the principal tools in the theoretical study of biological and synthetic macromolecules is the method of molecular dynamics (MD) simulations. This computational method calculates the time dependent behavior of a molecular system. It is based on Newton's second law, or the classical equation of motion, **F=ma**, where **F** is

the force exerted on the particle, **m** is its mass, and **a** is its acceleration. From the knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system.

Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. From this trajectory, the average values of properties can be determined. The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past. The most common ensembles used are NVE, NVT, and NPT, where thermostats and barostats are employed to keep temperature and pressure constant.

### 1.4.2 Force Field

The force field and the way it is implemented are two important factors contributing to the accuracy of an MD simulation. The force field is a classical molecular mechanics potential energy function that describes the structure and covalent connectivity of the molecules as well as the intermolecular interaction. Currently commonly used force fields for MD simulations of nucleic acids are AMBER [46] and CHARMM [47]. In this study, AMBER 94 [48] is used as the force field for all simulations on RNA tetraloops.

The importance of an accurate force field is always clear to all. After 1990s, it was realized that the implementation of the force field and the inclusion of solvent and counterions are just as critical to an accurate simulation as the force field parameters themselves [49]. Although it is possible to treat the solvent implicitly (i.e. with an effective dielectric constant that is dependent on distance or a more elaborate model to mimic solvent screening), now it is routine to run simulations of nucleic acids in explicit solvent with a proper treatment of the long-range electrostatic interactions, such as the particle mesh Ewald (PME [50, 51]) method.

Previous work determined both AMBER94 [48] and MMFF94 [52, 53, 54] are appropriate force field that can be used to accurately simulate the behavior of cyclopentane ring. However, for simulating the behavior of ROMP PNB MMFF94 appeared to be a more accurate force field [40]. Therefore in our study MMFF94 is used as the force field for all simulations on ROMP PNBDC.

### 1.4.3 Wide Angle X-ray Diffraction

Wide angle X-ray diffraction (WAXD) is a potential method to characterize the helix-kink type of polymer behavior. The normal diffraction pattern for an amorphous polymer is a broad peak, called an "amorphous halo". When order appears in the backbone conformation of the polymer, the amorphous halo will be split into two distinct peaks, one for the intermolecular interactions (low angle) and the other for intramolecular interactions (high angle).

Molecular simulations of non-ROMP PNB were found to produce WAXD spectra which matched that of the experimental data (also shown in Figure 1.4) [55]. This WAXD pattern can be explained with a helix-kink conformation which was imposed in the simulation. The model shows a split between the intermolecular (low angle) and intramolecular (high angle) interactions from two distinct patterns of order.

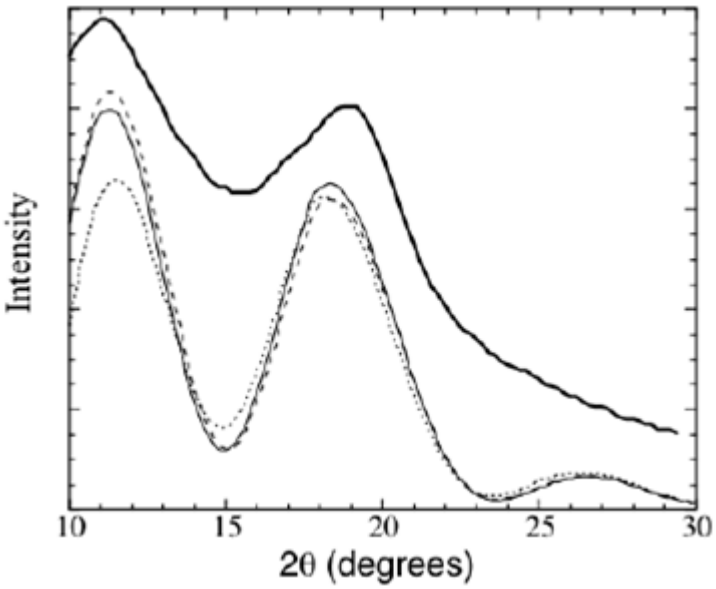



Figure 1.4: WAXD spectra for PNB. The dark solid line represents the experimental WAXD pattern for PNB, the dashed line represents a four chain model of PNB, the solid line represents a six chain model of PNB, and the dotted line represents the WAXD pattern for a poly(ethylene)-PNB 50:50 random copolymer [55].

## *1.5 Objectives and Hypotheses*

This work is organized in two parts. Part I (Chapter II – IV): studies thermodynamics of standard tetraloops and d2 tetraloops and the folding events of standard tetraloops. Part II (Chapter V): studies of the conformation of ROMP PNBDC_UD and the associated WAXD patterns.

### 1.5.1 Part I: RNA tetraloops

RNA tetraloops have high thermodynamic stability and plays an important role in a lot of important biological functions [14, 15, 29, 30, 31]. The deletion mode of DevLS (d2 motif) is observed abundantly in 23S rRNA of HM LSU by Williams and co-workers[32]. Additionally, Gutell and coworkers [56] have also found many examples of structure similar to the d2-tetraloop, what they call the Lonepair Triloop (LPTL) in 5S, 16S, and 23S rRNA. In comparison to standard tetraloops, due to the missing j+2 residue, d2 tetraloops has less intra-loop hydrogen bonding interactions and less base stacking in the loop area and therefore are considered a highly strained structures. Therefore we hypothesize that something about the d2 tetraloop must stabilze its structure.

Half of these d2 tetraloops (Williams) or Lonepair Triloops (Gutell) are in close association with proteins. The majority of them are positioned at sites in rRNAs that are conserved in three phylogenetic domains; a few of these occur in regions of rRNA associated with ribosomal function, including the presumed site of peptidyl transferase activity in the 23S rRNA. Because they are so conserved in various organisms and associated with such important functions it is important to understand structure.

The overall objective of our work on RNA tetraloops is to identify the stabilizing factor of the structure of d2 tetraloops and to probe the folding mechanism of the standard tetraloop through MD simulation. Our broad goals are to determine the reasons for the high abundance of the d2 tetraloop so as to better understand its varied functions, as well as to investigate whether d2 tetraloop is an intermediate during the folding of the standard tetraloop.

**Specific Aim 1. Investigate the role of stem defects in the d2 tetraloop.** The d2 tetraloop has less base pairing and stacking in the loop compared to the standard tetraloop, and is therefore considered less stable. In addition, the d2 tetraloop always occurs with some stem distortion. These stem defects are believed to contribute to the stability of the d2 tetraloop by providing some conformational relaxation in the stem. We will perform MD simulation on both the d2 tetraloop and perfect d2 tetraloop, which is a d2 tetraloop with the stem defect removed, to investigate the role of the stem defect. The following two hypotheses will be tested:

- *Hypothesis I: A perfect d2 tetraloop is not stable.*
- *Hypothesis II: Stem defects stabilize the d2-tetraloop with insertion.*

**Specific Aim 2. Investigate the folding pathway of the standard tetraloop.** We believe that another reason for the high abundance of the d2 tetraloop in the database is that the d2 tetraloop is an intermediate state kinetically trapped during the folding of the standard tetraloop. We will conduct high temperature unfolding simulation to investigate the folding of standard tetraloop. The following hypothesis will be tested:

- *Hypothesis III: A d2 tetraloop with insertion is a kinetically trapped intermediate in the folding of a standard tetraloop.*

### 1.5.2 Part II: ROMP PNBDC_UD

Helical polymers tend to form elongated structures, and such structures have

good catalyst site accessibility when used as a catalyst substrate. A wide angle X-ray scattering experiment conducted by Weck group generated a WAXD pattern with a split in the amorphous halo that is indicative of helical conformations.

Previous simulations of the ROMP PNB, which is being investigated as a catalyst substrate, did not show significant evidence of helical conformations. Our goal is to investigate the role of a backbone dicarboximide ring and a bulky undecyl side group in making ROMP PNB adopt a helical conformation and produce observed WAXD pattern. The overall objective of our work on ROMP poly(norbornene) is to identify the conformation of both ROMP PNBDC and ROMP PNBDC_UD that may reproduce this WAXD pattern.

**Specific Aim 3. Identify a helical conformation in ROMP PNBDC and PNBDC_UD that reproduces the observed WAXD pattern.** Studies have shown that ROMP PNB forms a random coil due to the lack of rigidity of cyclopentane ring, despite the fact that other polymers with alternating non-rotatable bonds and bulky side-groups (e.g. non-ROMP PNB) do form helices. To decrease the flexibility of the cyclopentane ring, a second dicarboximide ring will be fused to the cyclopentane ring to make it non-rotatable. This work is also aimed at investigating the role of the undecyl bulky side group by simulating both ROMP PNBDC and ROMP PNBDC_UD and performing bulk simulations on the identified configuration of ROMP PNBDC_UD that forms a helical structure to reproduce the experimentally generated WAXD patterns. The following hypothesis will be tested:

*Hypothesis IV: ROMP PNBDC_UD forms ordered helices in the bulk similar to non-ROMP PNB.*

In order to test hypothesis IV, single molecule MD simulations will be performed to determine the right combination of backbone torsion angles that enable

the helical conformation of ROMP PNBDC. A four chain model will then be built to perform bulk simulation. After bulk simulation, WAXD will be calculated and compared with experimental data.

## *1.6 References*


1. Brenner S., Jacob F., Meselson M.. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 1961, 190:576-581.
2. Crick F. The genetic code. *Cold Spring Harbor Symp Quant Biol*. 1965, 31.
3. Cech T. Structure and mechanisms of the large catalytic RNAs: group I and group II introns and ribonuclease P. In : *Gesteland R., Atkins J. (eds), The RNA World*, CSHL Press 1993, 239-269.
4. Gesteland R., Cech T., Atkins JF (eds). *The RNA World*. 1999; CSHL Press.
5. http://en.wikipedia.org/wiki/File:RNA_chemical_structure.GIF
6. http://en.wikipedia.org/wiki/File:Nucleotides_1.svg
7. Saenger, W. (1984) Principles of nucleic acid structure.).xx, 556 p., Springer-Verlag,New York.
8. Chworos, A., Severcan, I., Koyfman, A. Y., Weinkam, P., Oroudjev, E., Hansma, H. G. and Jaeger, L. Building programmable jigsaw puzzles with RNA. *Science* 2004, 306: 2068-2072.
9. Moore, P. B. Structural motifs in RNA. *Annu. Rev. Biochem*. 1999, 68: 287-300.
10. Leontis, N. B., & Westhof, E. Analysis of RNA motifs. *Curr Opin Struct Biol* 2003. 13: 300-308.
11. Bourdeau, V., Ferbeyre, G., Pageau, M., Paquin, B., Cedergren, R. The distribution of RNA motifs in natural sequences. *Nucleic Acids Res* 1999. 27: 4457-4467.
12. Woese, C. R., Gutell, R., Gupta, R. and Noller, H. F. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol. Rev*.

1983. 47: 621-669.

13. Woese, C. R., Winder, S. and Gutell R. R. Architecture of ribosomal-RNA-constraints on the sequence of tetraloops. *Proc. Natl. Acad. Sci., U.S.A.* 1990. 87: 8467-8471.
14. Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., d'Aubenton-Carafa, Y., Uhlenbeck, O. C., Tinoco, I., Jr., Brody, E. N. and Gold, L. CUUCGG hairpins:Extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci., U.S.A.* 1988. 85: 1364-1368.
15. Jaeger, L., Michel, F. and Westhof, E. Involvement of a GNRA tetraloop in long-range tertiary interactions. *J. Mol. Biol.* 1994. 236: 1271-1276.
16. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. and Pardi, A. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.* 1996. 264: 968-980.
17. Butcher, S. E., Diechmann, T. and Feigon, J. Solution structure of a GAAA tetraloop receptor RNA. *EMBO J.* 1997. 16: 7470-7499.
18. Jucker, F. M. and Pardi, A. GNRA tetraloops make a U-turn. *RNA*. 1995. 1: 219-222.
19. Correll, C. C. and Swinger, K. Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 Å resolution. *RNA*. 2003. 9: 355-363.
20. Cheong, C. Varani, G. and Tinoco, I., Jr. Solution structure of an unusually stable RNA hairpin, 5'ggac(UUCG)GUCC. *Nature* 1990. 346: 680-682.
21. Akke, M., Fiala, R., Jiang, F., Patel, D. and Palmer, A. G., 3rd. Base dynamics in a UUCG tetraloop RNA hairpin characterized by 15N spin relaxation: Correlations with structure and stability. *RNA* 1997. 3: 702-709.
22. Ennifar, E., Nikulin, A., Tishchenko, S., Serganov, A., Nevskaya, N., Garber, M., Ehresmann, B., Ehresmann, C., Nikonov, S. and Dumas, P. The crystal structure of UUCG tetraloop. *J. mol. Biol.* 2000. 304: 35-42.

23. Williams, D. J. and Hall, K. B. Unrestrained stochastic dynamics simulations of the UUCG tetraloop using an implicit solvation model. *Biophys. J.* 1999. 76: 3192-3205.

24. Allain, F. H. and Varani, G. Structure of the the p1helix from group I self-splicing introns. *J. Mol. Biol.* 1995. 250: 333-353.

25. Jucker, F. M. and Pardi, A. Solution structure of the CUUG hairpin loop – a novel RNA tetraloop motif. *Biochem.* 1995. 34: 14416-14427.

26. Baumruk, V., Gouyette, C., Huynh-Dinh, T., Sun, J. S. and Ghomi, M. Comparison between CUUG and UUCG tetraloops: Thermodynamic stability and structural features and analyzed by uv absorption and vibrational spectroscopy. *Nucleic Acids Res.* 2001. 29: 4089-4096.

27. Varani, G., Cheong, C. Tinoco, I., Jr. and Wimberly, B. Structure of an unusually stable RNA hairpin – conformation and dynamics of an RNA internal loop. *Biochem*. 1991. 30: 3280-3289.

28. Antao, V. P. and Tinoco, I., Jr. Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res.* 1992. 20:819-824.

29. Selinger, D., Liao, X. and Wise, J. A. Functional interchangeability of the structurally similar tetranucleotide loops GAAA and UUCG in fission yeast signal recognition particle RNA. *Proc. Natl. Acad. Sci., U.S.A.* 1993. 90: 5409-5413.

30. Michel, F. and Westhof, E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* 1990. 216: 585-610.

31. Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D. and Williamson, J. R., Conformation of the tar RNA-arginine complex by NMR spectroscopy. *Science* 1992. 257: 76-80.

32. Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.H., Srinivasan, A.R., and Schneider, B. (1992) The nucleic-acid database – a comprehensive relational database of 3-dimensional structures of nucleic-acids. *Biophysical Journal*, 63**,** 751-759.

33. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) Protein data bank – computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, 112**,** 535-542.
34. Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science*, 289**,** 905-20.
35. Klein, D.J., Schmeing, T.M., Moore, P.B., and Steitz, T.A. (2001) The kink-turn: a new RNA secondary structure motif. *Embo J*, 20**,** 4214-21.
36. Selmer, M., Dunham, C.M., Murphy, F.V.t., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R., and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, 313**,** 1935-42.
37. Juneau, K., Podell, E., Harrington, D.J., and Cech, T.R. (2001) Structural basis of the enhanced stability of a mutant ribozyme domain and a detailed view of RNA--solvent interactions. *Structure*, 9**,** 221-31.
38. Hsiao, C., Mohan, S., Hershkovitz, E., Tannenbaum, A., & Williams, L. D. (2006). Single nucleotide RNA choreography. *Nucl. Acids Res*. 34 (5), 1481–1491.
39. Clark, J. H. and Macquarrie, D. J., Kirk-Othmer Encyclopedia of Chemical Technology Online, vol. 5, ch. Supported Catalysts, pp. 322–344. Wiley-Interscience, 2002.
40. Swann, A. T., *Characterization of polymer-supported homogeneous catalysts by molecular modeling.* PhD thesis, Georgia Institute of Technology, 2008.
41. Sattigeri, J. A., Shiau, C. W., Hsu, C. C., Yeh, F. F., Liou, S., Jin, B. Y., and Luh, T. Y., "Remarkable enhancement of second-order nonlinear optical properties of polynorbornenes having pendant chromophores. use of hyper-rayleigh scattering to estimate the tacticity of rigid rod polymers," *Journal of the American ChemicalSociety*, vol. 121, pp. 1607–1608, Feb. 1999.
42. Lin, W. Y., Murugesh, M. G., Sudhakar, S., Yang, H. C., Tai, H. C., Chang, C. S., Liu, Y. H., Wang, Y., Chen, I. W. P., Chen, C. H., and Luh, T. Y., "On the rigidity

of polynorbornenes with dipolar pendant groups," *Chemistry-A European Journal,* vol. 12, pp. 324–330, 2006.

43. Diaz, K., Vargas, J., Del Castillo, L. F., Tlenkopatchev, M. A., and Aguilar-Vega, M., "Polynorbornene dicarboximides with cyclic pendant groups: synthesis and gas transport properties," *Macromolecular Chemistry and Physics*, vol. 206, pp. 2316–2322, 2005.
44. Contreras, A. P., Cerda, A. M., and Tlenkopatchev, M. A., "Synthesis of hightg polymers by ring-opening metathesis polymerization of n-cycloalkylnorbornene dicarboximide," Macromolecular Chemistry and Physics, vol. 203, pp. 1811–1818, 2002.
45. Hino, T., Inoue, N., and Endo, T., "Ring-opening metathesis copolymerization behaviors of cyclooctene and norbornene bearing a five- or six-membered ring cyclic carbonate," *Journal of Polymer Science Part A-Polymer Chemistry,* vol. 43, p. 6599, Dec. 2005.
46. D.A. Case, T.E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. Woods. The Amber biomolecular simulation programs. *J. Computat. Chem.* 26, 1668-1688 (2005).
47. MacKerell, A.D., Jr. "Atomistic Models and Force Fields" in Computational Biochemistry and Biophysics, O.M. Becker, A.D. MacKerell, Jr., B. Roux and M.Watanabe, Eds., Marcel Dekker, Inc. New York, 2001, p. 7-38.
48. Cornell, W. D., Cieplak P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., & Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117 (19): 5179-5197, 1995
49. Thomas E. Cheatham, III, Peter A. Kollman. Molecular Dynamics Simulations of Nucleic Acids. *Annu. Rev. Phys. Chem.* 2000. 51: 435-471.
50. Darden TA, York DM, Pedersen LG. Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems. *J. Chem. Phys*. 1993. 98:10089-10092.
51. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. *J. Chem.*

*Phys.* 1995. 103: 8577-8593.

52. Thomas A. Halgren. Merck molecular force field. I. Basis, form, scope, parameterization and performance of MMFF94. *J. Comput. Chem.* 17: 490-519, 1996.
53. Thomas A. Halgren. Merck molecular force field. II. MMFF94 van der waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* 17: 520-552, 1996.
54. Thomas A. Halgren. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem*. 17: 553-586, 1996.
55. Ahmed, S., Ludovice, P. J., and Kohl, P., "Microstructure of 2,3 erythro di-isotactic polynorbornene," *Computational and Theoretical Polymer Science,* vol. 10, pp. 221–233, 2000.
56. Lee,J.C., Cannone,J.J. and Gutell,R.R. (2003) The lonepair triloop: a new motif in RNA structure. J. Mol. Biol., 325, 65–83.

# CHAPTER II

## STANDARD TETRALOOP AND

## PERFECT D2 TETRALOOP

This chapter presents the structural difference and the stability comparison of RNA standard tetraloop and perfect d2 tetraloop determined by Molecular Dynamics simulation.

## *2.1 Introduction*

### 2.1.1. Background

*2.1.1.1 RNA Deviations of Local Structure (DevLS)*

Tetraloops are known to be the simplest, smallest and most frequent RNA motif. Williams and Coworkers [1] show that even at this most elementary level of organization, RNA tolerates variation in conformation, topology, and molecular interactions. The variation is described by four distinct modes, which are insertions, deletions, strand clips and 3-2 switches, collectively called DevLS (Deviations of Local Structure, Figure 2.1). The numbers in the figure indicate the covalent ordering of the residues along the polynucleotide strand. These four DevLS arise from common enabling factors, which operate at the single nucleotide level. These common factors are the high RNA backbone length per residue (six bonds separate adjacent residues) and numerous torsional degrees of freedom of RNA nucleotides.

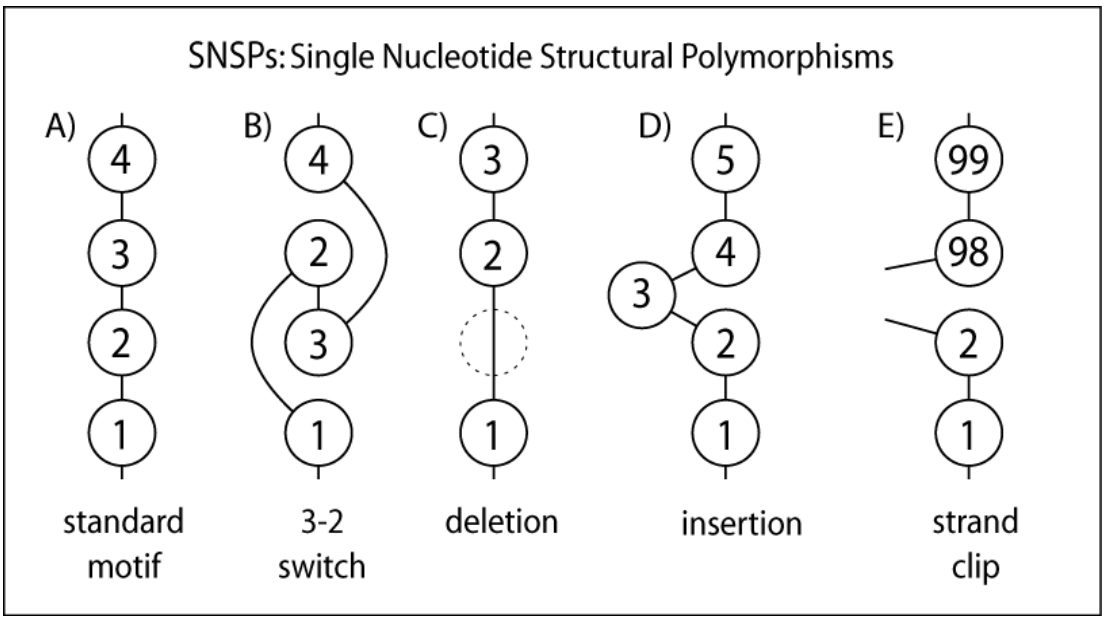



Figure 2.1: RNA DevLS. A) A generic RNA motif is represented schematically by four circles, which symbolize four residues. B) In a motif with 3-2 switch, the positions of two bases, of residues 3 and 2 in the figure, are interchanged. The backbone linkage is maintained. C) In a deleted motif, a residue is omitted (dashed line). D) In an inserted motif, a residue is added. E) In a strand clipped motif one or more residues is contributed from a remote region of the primary sequence. An insertion, if extensive enough can be equivalent to a strand clip [1].

*2.1.1.2 Standard Tetraloop and d2 Tetraloop*

In the tetraloop family, the standard tetraloop is the standard motif or local structure described earlier in RNA DevLS. The position nomenclature and the structure of a standard tetraloop are exhibited in Figure 2.2. The positions of the four nucleotides in the loop area are j-1, j, j+1, and j+2 respectively. Thermodynamic studies have shown that frequently occurring RNA standard tetraloops (the UNCG, the GNRA, and the CUUG tetraloop sequences, where N can be any nucleotide and R is either G or A) are very stable [2, 3, 4]. In the loop area, there are two characterizing hydrogen bonding

interactions, one between residue j-1 and residue j+1 and the other between residue j-1 and residue j+2. There are also base stacking interactions among residues at position j, j+1, and j+2. The stem of a standard tetraloop also has perfect base pairing and stacking. All of these interactions are reasons for the stability of a standard tetraloop.

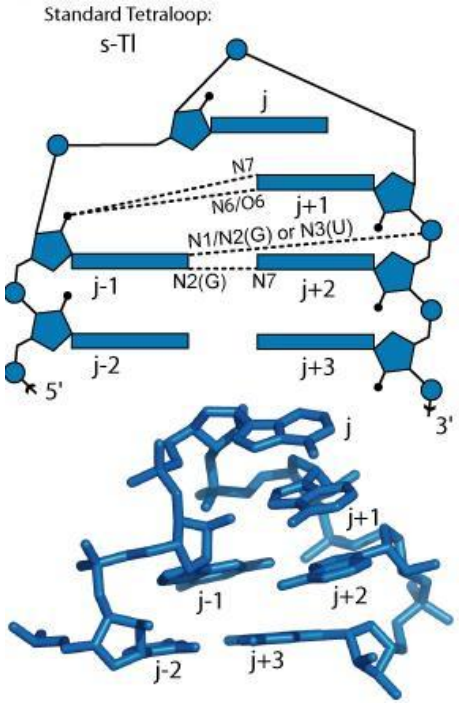



Figure 2.2: Residue position nomenclature and secondary structure of a standard tetraloop [1].

In contrast, the d2 tetraloop is a deviation from the standard motif, the standard tetraloop in this case, and belongs to the deletion family of DevLS [1]. In the d2 tetraloop, residue j+2 of the standard tetraloop is absent (Figure 2.3). Residue j+3 of the standard tetraloop becomes j+2 of the d2-tetraloop. The deletion of residue j+2 in the loop area does not appreciably change the positions of the remaining backbone atoms of these tetraloops. However Williams and coworkers found that the deletion at position 2 is always correlated with adjacent helical distortions such as insertions, clipping, or

base pairing disruption in the stem. In other words, a perfect d2 tetraloop, referring to a d2 tetraloop without any defects in the stem, has not been observed.

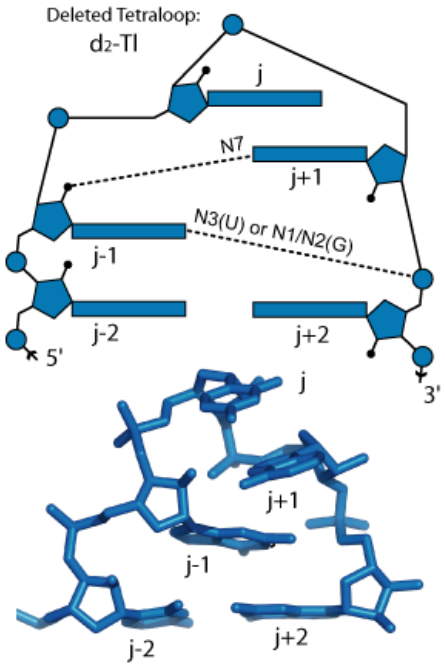



Figure 2.3: Residue position nomenclature and secondary structure of a d2 tetraloop [1].

*2.1.1.3 Rarity of Perfect d2 Tetraloop*

As discussed in Chapter I, the HM 23S rRNA (1JJ2) is our test 'database', hereafter referred to as the database. The crystal structure of the large ribosomal subunit from *Haloarcula marismortui* has been determined to high resolution by Steitz and Moore [5, 6]. At 2.4 Å resolution, the atomic positions of the vast majority of the 23S rRNA of HM LSU are well-characterized and are more accurately determined than any other large RNA complex. The HM 23S rRNA, with over 2500 residues, constitutes a large database with a large variety of RNA conformation and interactions.

Williams and coworkers observe 40 structurally related tetraloops through multi-resolution analysis from the database (Figure 2.4) [1]. Out of the 40, 21 are

standard tetraloops. Surprisingly 10 are d2-tetraloops which represent an abundant existence in the database. Again no perfect d2 tetraloop with nice base pairing and stacking in the stem like standard tetraloop has been observed. This suggests that stem defects in d2 tetraloop may play an important role in stabilizing this conformation.

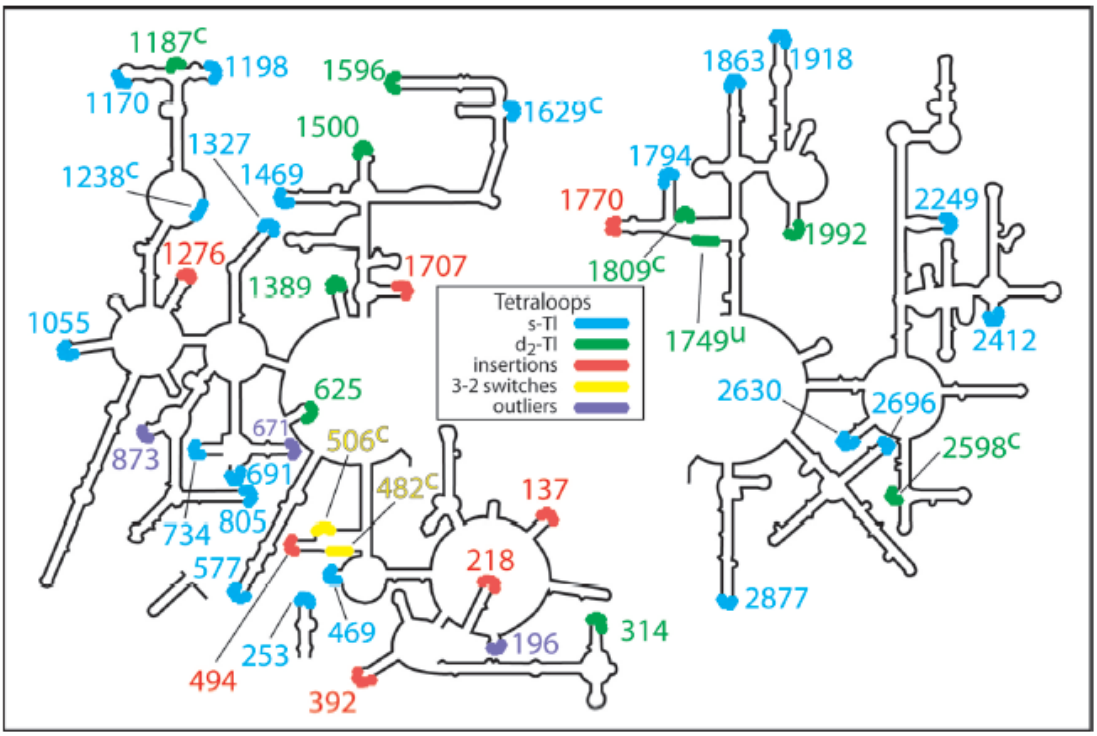



Figure 2.4: Secondary structure of the HM 23S rRNA (1JJ2). Tetraloop locations and type are indicated by color. Superscripted c's indicate strand clipped tetraloops [1].

### 2.1.2 Motivation

In the loop area of the standard tetraloop (Figure 2.2), the hydroxyl group on the sugar of residue j-1 forms a hydrogen bond interaction with a proton acceptor on the base of residue j+1. In addition, the base of residue j-1 can form hydrogen bonding interaction with either the base or the backbone atom of residue j+2. Besides hydrogen bonding interactions, base stacking is also an important interaction for keeping the loop structure stable. For example, the base of residue j+1 forms stacking with the base of

residue j+2. All these intra-loop interactions are partially responsible for such a stable structure of the standard tetraloop.

In comparison to standard tetraloops, due to the missing residue j+2, d2 tetraloops have less intra-loop hydrogen bonding interactions and less base stacking in the loop area (Figure 2.3). There are still hydrogen bonding interactions in the loop region of the d2 tetraloop, such as the one between the hydroxyl group on the sugar of residue j-1 and the base of residue j+1, and the one between the base of residue j-1 and the backbone atom of residue j+2. However, because residue j+2 is shifted down to the j+3 position in the standard tetraloop and becomes part of the closing base pair of the d2 tetraloop, the hydrogen bonding interaction between the base of residue j-1 and the base of residue of j+2 that exists in the standard tetraloop is absent in the d2 tetraloop. Moreover, the base stacking interaction between the base of residue j+1 and the base of residue j+2 that exists in the standard tetraloop is also absent in the d2 tetraloop. As a result, the d2 tetraloop is considered to be more strained and less stable than standard tetraloop.

Williams and coworkers have found that the deletion at position j+2 is always correlated with adjacent helical distortion. Unlike standard tetraloops, which have a healthy stem with perfect base pairing and stacking, d2-tetraloops always appear with some defects in the stem. In comparison to standard tetraloops, due to the missing residue j+2, d2-tetraloops form an unhealthy stem with adjacent helical imperfection such as insertion (314, 625, 1387, 1992), clipping (1187, 1809, 2598), base pairing disruption in the stem (1500, 1596) and unhinging (1749) (Figure 2.4) [1]. Therefore we raise the question: would a perfect d2 tetraloop with no defects in the stem be a stable structure?

The method of Molecular Dynamics (MD) simulations has long been used as one of the principal tools in the theoretical study of biological molecules such as proteins and nucleic acids. MD simulations have provided detailed information on the fluctuations and conformational changes of proteins and nucleic acids. These methods are now routinely used to investigate the structure, dynamics and thermodynamics of

biological molecules and their complexes. Here MD simulations are performed on both a sample structure of the perfect d2-tetraloop and a sample structure of the standard tetraloop to test whether the perfect d2 tetraloop would be stable without its stem defects.

### 2.1.3 Objectives

The objective here is to find out whether the perfect d2 tetraloop with no defects in the stem will remain stable by running molecular dynamics simulations on both the perfect d2 tetraloop (artificially built) and the standard tetraloop (from the database) and comparing these two simulations.

*Hypothesis I. The perfect d2 tetraloop is not stable relative to the standard tetraloop.*

To test Hypothesis I, a sample structure of the d2 tetraloop with certain stem defects and a sample structure of the standard tetraloop can be taken from the database. The sample structure of d2 tetraloop can then be modified by removing the defects in the stem to artificially build a perfect d2 tetraloop with a perfect healthy stem. Afterwards MD simulations can be performed on both the standard tetraloop and the perfect d2 tetraloop with no stem defects. Simulation results can be compared between the perfect d2 tetraloop and the standard tetraloop to see if the structure of a perfect d2 tetraloop is stable. If simulation results showed the perfect d2 tetraloop to be unstable, future study on the d2 tetraloop with defects in the stem would be focused on how stem defects stabilize the d2 tetraloop.

## *2.2 Methods*

### 2.2.1 Modification of a Sample d2 Tetraloop with Stem Defects

A sample structure of the d2 tetraloop with defects in the stem was chosen from the database to test hypothesis I. This sample structure was labeled as the 314-d2-tetraloop. The 314-d2-tetraloop is a d2-tetraloop with a single residue insertion in the stem, an example of stem defects. The inserted residue is located between residue j+3 and j+4 on the 3' side of the stem. The number (314) represents the nucleotide number in HM 23S rRNA secondary structure and is also the first residue number (at position j-1) in the loop. The reason that this specific sample structure was chosen is that it is the simplest d2-tetraloop in the database with only one insertion residue in the stem. Since the d2 tetraloop with stem defects needs to be simulated in the next chapter, choosing this smaller structure will make the simulations computationally more expedient.

The 314-d2-tetraloop has a sequence of CUU(GGA)A(C)AG and is a d2-tetraloop with one residue insertion (318C) in the stem (Figure 2.5). The first parentheses in the sequence are used to indicate that the nucleotides inside the bracket are the residues (314G, 315G and 316A) inside the loop. The second parentheses is used to indicate that the nucleotide inside the parentheses is the inserted residue (318C) in the stem. In order to test hypothesis I, the insertion residue 318C is artificially removed and the remaining two residues, residue 317A and residue 319A are covalently bonded. The d2 tetraloop is modified by removing its insertion residue in the stem and becomes a perfect d2 tetraloop with a healthy stem (no insertion residue in the stem). This perfect d2 tetraloop with sequence CUU(GGA)AAG is labelled as cutd2-tetraloop (Figure 2.5). If a perfect d2 tetraloop with no defects in the stem is indeed unstable, this cutd2-tetraloop would not be stable during the simulation and would eventually fall apart.

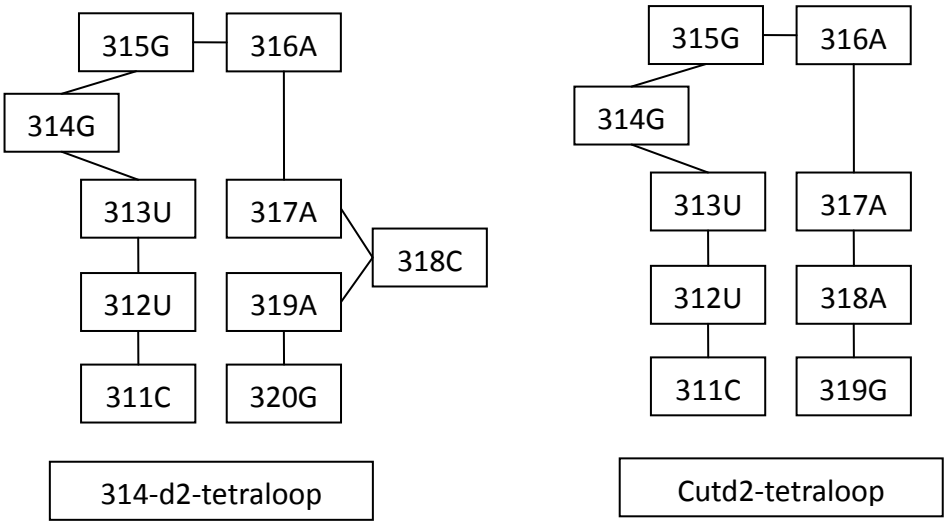



Figure 2.5: A rough sketch of the secondary structure of 314-d2-tetraloop (left) and cutd2-tetraloop (right).

### 2.2.2 Simulation Methods

*2.2.2.1 Molecular Dynamics (MD) Simulation*

Molecular Dynamics (MD) simulations are widely used to investigate questions of macromolecular structure and dynamics. There are many algorithms and simulation parameters (such as force field parameters, temperature controlling algorithms, etc.) available to choose when doing simulation. The simulation parameters and algorithms were tested by simulating a sample structure of the standard tetraloop, here labeled as 805-standard tetraloop, from HM 23S rRNA database. This 805-standard tetraloop has a sequence of GCC(GAAA)GGC (Figure 2.6). 805 standard tetraloop is a typical GNRA tetraloop, the structure of which has been determined to high resolution. By comparing our simulation results to experimental results, the simulation methods can be proven to be robust and can further be used for cutd2-tetraloop simulations and other tetraloops simulations in this study. A

comparison of simulations on the standard tetraloop (805-standard tetraloop) and the perfect d2 tetraloop (cutd2-tetraloop) was conducted.

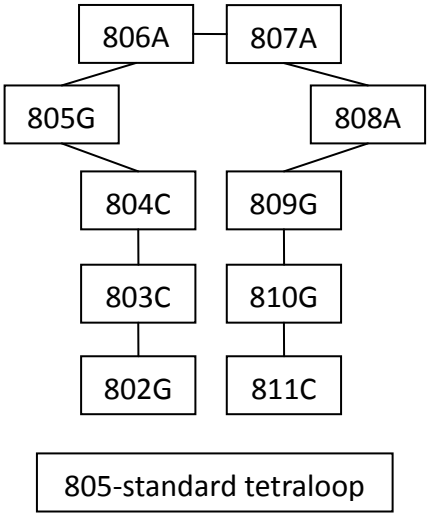



Figure 2.6: A schematic of the secondary structure of 805-standard tetraloop

*2.1.2.2 MD Simulation Protocol*

Using AMBER94 as the force field with the NAMD program [8], a box of 1735 TIP3P water molecules [9] was added around the RNA 805-standard tetraloop to a depth of 12 Å from the edge of the RNA 805-standard tetraloop. Also, 12 $Na^+$ and 12 $Cl^-$ counterions were placed within the water box to ionize the system. Nine extra $Na^+$ were also added to neutralize the system. Both the solvation and the ionization were done through VMD psfgen package [10].

Simulations with a time step of 2 femtosecond (fs) were carried out using the Particle Mesh Eward (PME) technique for long range electrostatic interactions calculations and repeating boundary conditions in a box of $(40\ Å)^3$ with 12 Å nonbonded cutoff [11, 12]. Equilibration started with 5,000 steps of minimization followed by 700 ps of heating under MD, with atomic positions of RNA tetraloop molecules fixed. The system was slowly heated from 10 K to 310 K with a temperature reassigning frequency of every 10,000 steps.

The production MD runs were carried out with constant pressure boundary conditions (relaxation time 500 fs). A constant temperature of 310 K was applied using Berendsen Algorithm [14] with a coupling constant of 500 fs. SHAKE constraints were applied to all hydrogens to eliminate X-H vibrations [13], which justified a longer simulation time step. Simulation runs were in duration of 10 nanoseconds (ns). Simulations were also carried out with Langevin piston for pressure control (damping coefficient of 5/ps) and Langevin Dynamics for temperature control [15] for comparison with Berendsen Algorithm.

## *2.3 Results and Discussions*

### 2.3.1 Berendsen Temperature Coupling vs. Langevin Dynamics Temperature Control

The MD algorithm numerically integrates Newton's equations of motion. Since most experiments are usually conducted at constant temperature and constant pressure, which corresponds to the isothermic-isobaric ensemble, standard MD packages offer different approachess for maintaining constant temperature. Among these are periodic velocity rescaling and other methods that uniformly modify the atomic velocities. This work initially used the Berendsen temperature coupling [14], which is also a velocity rescaling method, as the temperature controller. It is a method

in which forces are added or reduced to simulate the coupling of the system to a heat bath of a specific temperature.

Harvey and coworkers [16] found that periodic velocity rescaling leads to an unexpected problem: a gradual bleeding of kinetic energy from high frequency motions such as bond stretching and angle bending into low frequency motions. This represents a violation of the principal of equipartition of energy, which requires that each degree of freedom has the same mean kinetic energy. They showed a net conversion of kinetic energy from the internal degree of freedom to the translation of the center of mass. With repeated rescaling, there will be a gradual loss of vibrational kinetic energy and a corresponding increase in the translational kinetic energy. Therefore, almost all of the kinetic energy is converted into the net translation of and rotation about the center of mass, producing a system with almost no kinetic energy associated with the internal degrees of freedom. This phenomenon was called "the flying ice cube" and is an artifact of Berendsen temperature controller.

Berendsen temperature control method is a velocity rescaling approach that maintains constant temperature by coupling to a temperature bath, which represents a gradual relaxation back toward the target temperature. Both the MD simulations of cutd2-tetraloop and 805-standard tetraloop using Berendsen temperature control have encountered the problem of "the flying ice cube" described in Harvey's paper. The center of mass translational motion was not periodically removed in both simulations. Because the center of mass translational energy could not couple back into the system, the center of mass translational energy grew with repeated rescaling. From the simulation window, the RNA molecule was observed to move faster and faster along one direction.

Harvey and coworkers pointed out three approaches to prevent the simulation artifacts caused by periodic velocity rescaling: 1). Maintaining temperature by reassigning velocities rather than rescaling. 2). Periodically removing motion of the center of mass. 3). Modifying algorithms that use velocity rescaling.

One way to fix the problem is to use langevin dynamics for temperature

control [15], where additional damping and random forces are introduced to the system. With Langevin Dynamics, a Langevin-like equation of motion is applied to each atom, with forces from the force field plus a frictional term plus a stochastic term. This is an efficient way to add or remove energy to every atom and thus regulate the temperature. After Langevin Dynamics was used to replace the Berendsen approach for temperature control, the 'flying ice cube' phenomenon was no longer observed. Therefore, Langevin Dynamics was used as the temperature controlling method for all simulations in this study.

### 2.3.2 Simulation Observations

#### *2.3.2.1 Simulation Results of Standard Tetraloop*

The purpose of simulating 805-standard tetraloop is threefold. First, a stable simulation of 805-standard tetraloop can show that the simulation methods (simulation parameters and algorithms) chosen are able to provide accurate results and therefore can be used for further studies. Second, a stable simulation of 805-standard tetraloop can also prove 10 residues is the appropriate length of structure. Third, the simulation results of the standard tetraloop can provide a benchmark for comparison to simulation results of the perfect d2 tetraloop.

Throughout 10 ns of the NPT MD simulation, the structure of 805-standard tetraloop remained structurally intact. The positions of backbone atoms did not change appreciably (RMSDs were calculated in the next section). Figure 2.7 highlighted the hydrogen bonding interactions observed during the simulation with dotted lines. There was perfect base pairing and base stacking in the stem (Figure 2.7) throughout the simulation. The two characteristic hydrogen bonding interactions in the loop region, the interaction between residue 805G at position j-1 and residue 807A at position j+1 and the interaction between residue 805G at j-1 and residue 808A at j+2, were conserved throughout the simulation (atom distances indicating the hydrogen bonding interactions

were calculated in the next section). There also appeared to be significant base stacking between 806A at j and 807A at j+1 as well as 807A at j+1 and 808A at j+2.

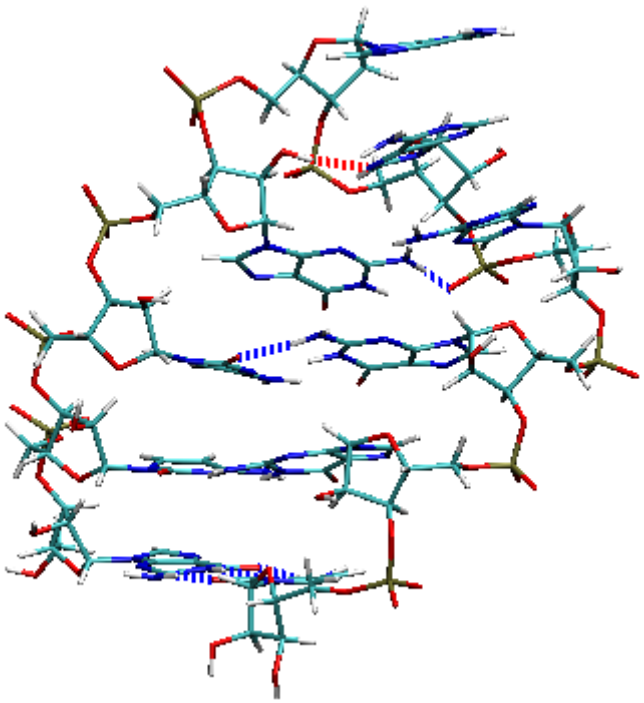


Figure 2.7: The structure of 805-standard tetraloop with hydrogen bond highlighted in dotted line at the end of 10ns simulation.

The snapshot of 805-standard tetraloop taken at the beginning of the simulation is placed side by side with the snapshot taken at the end of the simulation (Figure 2.8). The comparison showed that the structure of the standard tetraloop, 805-standard tetraloop in this case, was conserved during the simulation. An animation generated from series of snapshots at different time of the simulation (from start to end) showed that all the atoms were just vibrating slightly throughout the simulation.

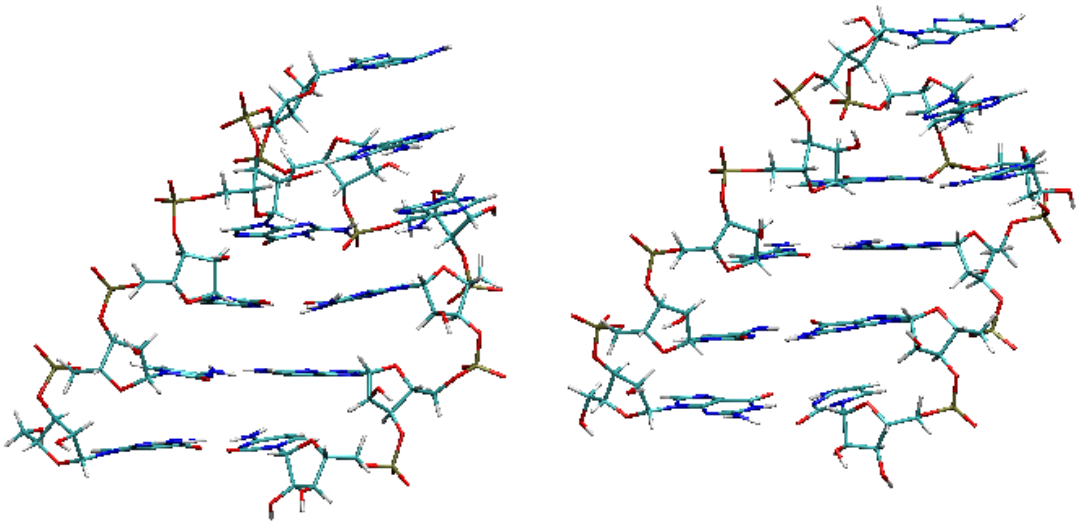


Figure 2.8: The snapshots taken from the MD simulation of 805-standard tetraloop, the left figure is the starting point, and the right figure is the end structure of 10ns of simulation.

*2.3.2.2 Simulation Results of the Perfect d2 Tetraloop*

The simulation of the perfect d2 tetraloop, cutd2-tetraloop in this case, exhibited an unstable structure. Unlike 805-standard tetraloop, the positions of cutd2-tetraloop backbone atoms changed significantly during 10 ns of MD simulation (RMSDs are calculated in the next section). The loop structure deformed considerably by the end of the simulation (Figure 2.9). The stacking between base 315G at position j and base 316A at position j+1 in the loop that existed at the beginning of the simulation was lost by the end of the simulation. Neither of the two intra-loop hydrogen bonding interactions, the one between residue 314G at j-1 and residue 316A at j+1 and the one between residue 314G at j-1 and residue 317A at j+2, were conserved in the simulation. The loop became very loose and unstructured.

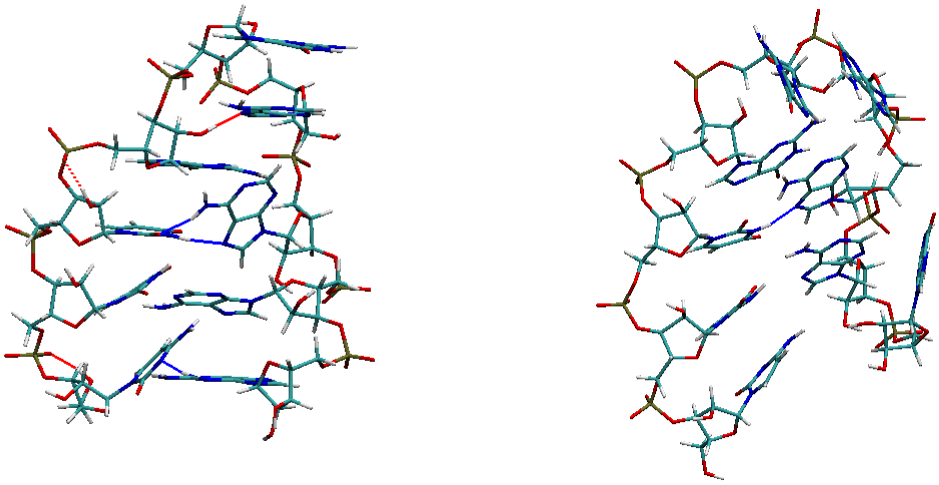


Figure 2.9: The snapshots from the MD simulation of cutd2-tetraloop, the left figure is the starting point, and the right figure is the structure at the end of the simulation.

The structure of tetraloops in this work has three base pairs in the stem. The bottom base pair at the end of the stem is referred to as the terminus base pair, e.g. 311C – 319G in Figure 2.5 (cutd2-tetraloop) and 802G – 811C in Figure 2.6 (805-standard tetraloop). The base pair close to the loop region at the top of the stem is referred to as the closing base pair, e.g. 313U – 317A in Figure 2.5 and 804C – 809G in Figure 2.6. The base pair between the terminus base pair and the closing base pair is referred to as the central base pair, e.g. 312U – 318A in Figure 2.5 and 803C – 810G in Figure 2.6.

In the simulation of cutd2-tetraloop, the hydrogen bonding interactions between the closing base pair 313U – 317A opened up during the first 2 ns but reformed for the rest of simulation time. The central base pair 313U – 317A could barely hold and

underwent significant atomistic changes. The terminus base pair 311C – 320G began to open soon after the simulation started, with the base of residue 320G pointing outward and the base of residue 311C trying to stack with the adjacent base of residue 312U. The stem was observed to unravel itself soon during the simulation.

The simulation of cutd2-tetraloop showed that a perfect d2 tetraloop with the insertion (stem defects) removed was very unstable. The structures of both the loop and the stem in cutd2-tetraloop changed considerably during the simulation compared to 805-standard tetraloop. By comparing the end structures of 805-standard tetraloop and cutd2-tetraloop after the simulation (Figure 2.10), we can see that 805-standard tetraloop maintained a loop and stem structure with perfect base pairing and stacking, while cutd2-tetraloop started to fall apart with the opening stem and deformed loop during the simulation.

At this point we conclude qualitatively that the perfect d2 tetraloop with no defects in the stem, cutd2-tetraloop with the insertion residue removed from its stem in this case, is not a stable structure compared to the standard tetraloop, 805-standard tetraloop simulated here. Therefore simulation observations of both the perfect d2 tetraloop and the standard tetraloop qualitatively validated hypothesis I. Quantitative analysis of simulation results was further conducted on both standard tetraloop and perfect d2 tetraloop in the following section.

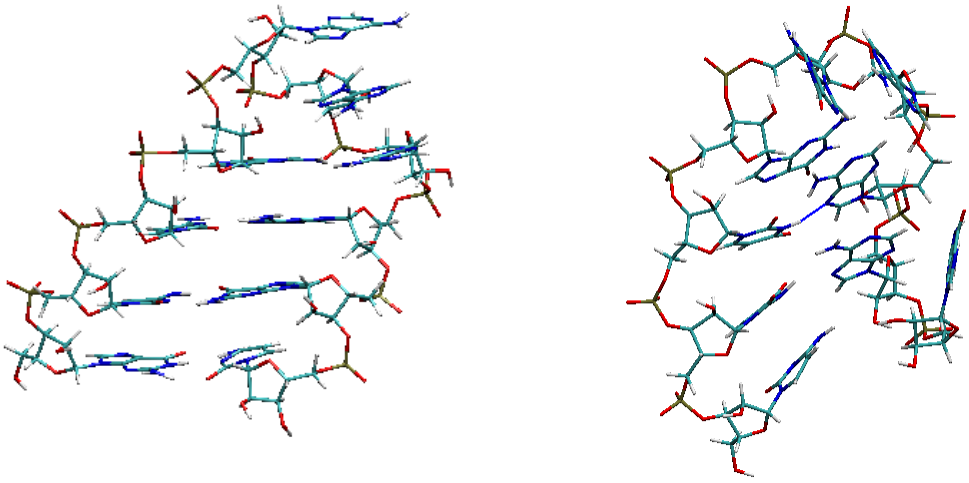


Figure 2.10: The end structures of 805 standard tetraloop (left) and cutd2-tetraloop (right) from the simulation.

### 2.3.3 Quantitative Comparisons of Simulation Results between Standard Tetraloop and Perfect d2 Tetraloop

The structural observations and comparisons from the molecular dynamics simulations on the standard tetraloop (805 standard tetraloop) and the perfect d2 tetraloop (cutd2-tetraloop) qualitatively indicated that the perfect d2 tetraloop is less stable than the standard tetraloop.

In this section, quantitative comparisons between the standard tetraloop (805 standard tetraloop) and the perfect d2 tetraloop (cutd2-tetraloop) were conducted. These included RMSD analysis, hydrogen bonding interactions tracking, and torsion angle analysis.

#### *2.3.3.1 RMSD Analysis*

Root mean square deviation (RMSD) values are commonly used to measure the structural similarity among biomolecule structures and were extensively used in this study for stability comparisons. Measuring RMSD values, in units of Ångström (Å), from simulation trajectories has also been frequently used as a quantitative way of examining the stability of a specific structure as well as comparing stabilities among different structures.

The RMSDs of all residues, loop region and stem region of both 805-standard tetraloop and cutd2-tetraloop were calculated relative to the starting structure and plotted on the same graph recording structural changes during the simulations (Figure 2.11, Figure 2.12, and Figure 2.13). The trajectories were superposed on the starting structure before the RMSDs were calculated.

All residue RMSD values of 805-standard tetraloop (Figure 2.11) showed that the structure of the standard tetraloop remained intact throughout the simulation. In contrast, RMSD values of cutd2-tetraloop were significantly higher showing that the structure of the perfect d2 tetraloop deviated significantly from its starting point. The comparison of all residue RMSDs calculation between the standard tetraloop and the perfect d2 tetraloop indicated that the perfect d2 with stem defects, cutd2-tetraloop with no insertion residue in this case, is not a stable structure.

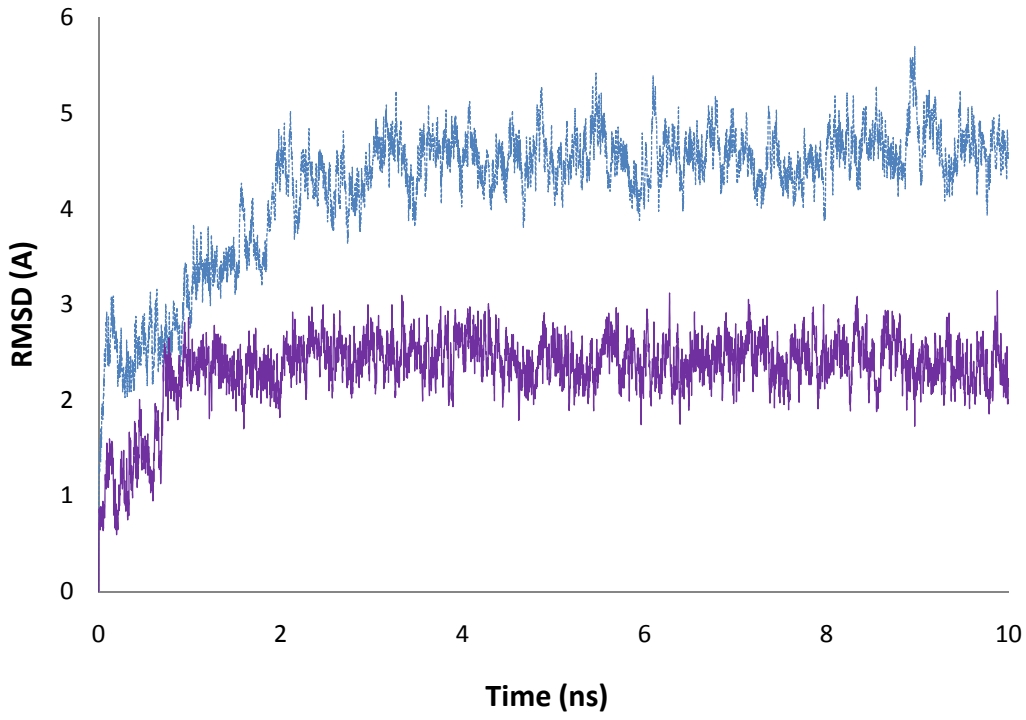



Figure 2.11: All residues RMSD calculation for 805-standard tetraloop (purple solid line) and cutd2-tetraloop (blue dotted line).

Figure 2.11 plotted all residues RMSD of both 805-standard tetraloop and cutd2-tetraloop over 10ns of simulation time. All residues RMSD of 805-standard tetraloop had a mean and a standard deviation of 2.33 Å and 0.38 Å, while all residues RMSDs of cutd2-tetraloop had a mean and a standard deviation of 4.24 Å and 0.71 Å. The significantly higher RMSDs of cutd2-tetraloop indicated that the perfect d2 tetraloop is less stable compared to the standard tetraloop.

Further investigating the structure by breaking it down into the loop region and the stem region allow us to see the difference in RMSDs changes over time between the standard tetraloop and the perfect d2 tetraloop more clearly. The loop region RMSD of 805-standard tetraloop has a mean and a standard deviation of 1.19 Å and 0.27 Å (Figure 2.12). In contrast, the loop region RMSD of cutd2-tetraloop has a mean and a standard deviation of 4.52 Å and 0.76 Å (Figure 2.12). Therefore, the

standard tetraloop appeared to have a more rigid loop compared to the stem (Figure 2.13). The loop region RMSD of cutd2-tetraloop showed that the loop of the perfect d2 tetraloop had a significant atomistic deviation from its starting point.

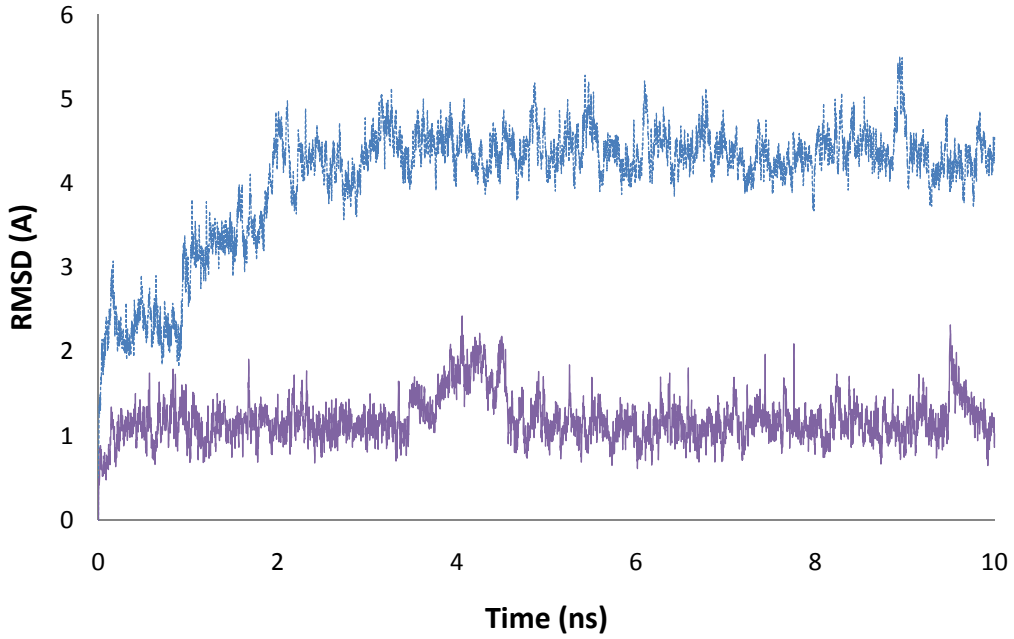



Figure 2.12: Loop region RMSD calculations for 805-standard tetraloop (purple solid line) and cutd2-tetraloop (blue dotted line).

The stem region RMSD of 805-standard tetraloop had a mean and a standard deviation of 1.58 Å and 0.32 Å (Figure 2.13). In contrast, the loop region RMSD of cutd2-tetraloop had a mean and a standard deviation of 4.07 Å and 0.71 Å (Figure 2.13). Therefore the stem region of the standard tetraloop appeared to be more flexible relative to the loop region. Again the stem region RMSD of cutd2-tetraloop showed that the stem of the perfect d2 tetraloop also had a significant atomistic deviation from its starting point. All residues, loop and stem RMSDs statistics were summarized in Table 3.1.

In summary, the comparison of all residues, loop and stem RMSDs

calculations of 805-standard tetraloop and cutd2-tetraloop showed that the structure of the perfect d2 tetraloop was not structurally stable relative to the d2 tetraloop.

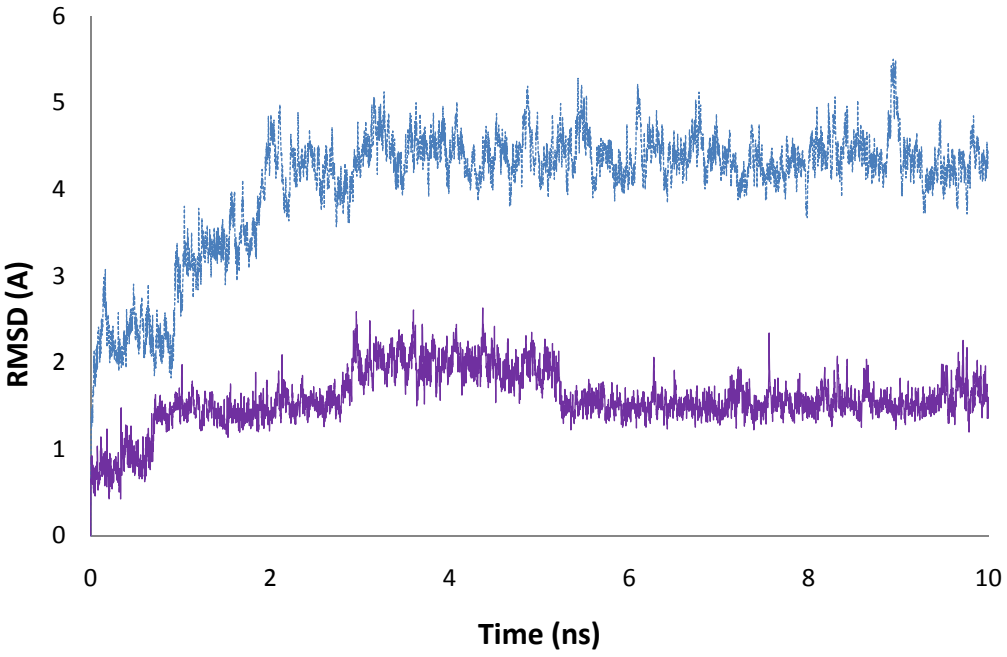



Figure 2.13: Stem region RMSD calculations for 805-standard tetraloop (purple solid line) and cutd2-tetraloop (blue dotted line).

To further investigate how the stem of the perfect d2 tetraloop without any defects in the stem (cutd2-tetraloop with insertion residue 318C removed) started to unravel during 10 ns of simulation, the RMSDs of the terminus base pair, the central base pair and the closing base pair of the perfect d2 tetraloop (cutd2-tetraloop) were calculated over 10ns of simulation time and were plotted on the same graph with the RMSDs of the terminus base pair, the central base pair and the closing base pair of the standard tetraloop (805-standard tetraloop) for comparison (Figure 2.14, Figure 2.15, and Figure2.16).

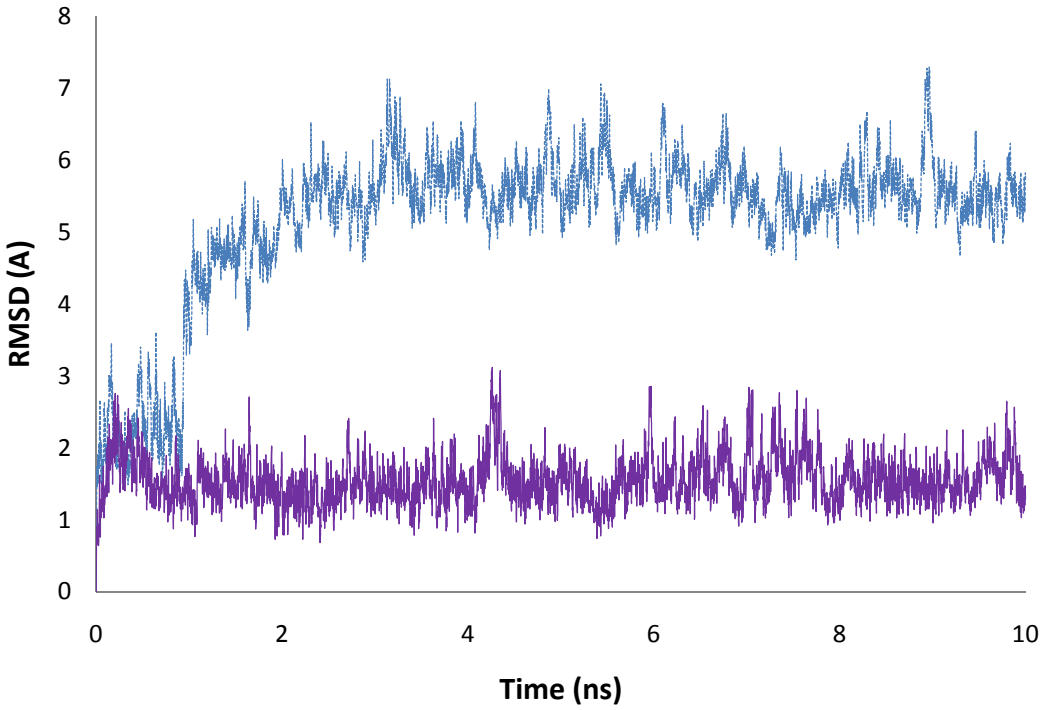



Figure 2.14: The terminus base pair RMSDs calculated over 10 ns of simulation for both the standard tetraloop (805-standard tetraloop in solid purple line) and the perfect d2 tetraloop (cutd2-tetraloop in dotted blue line).

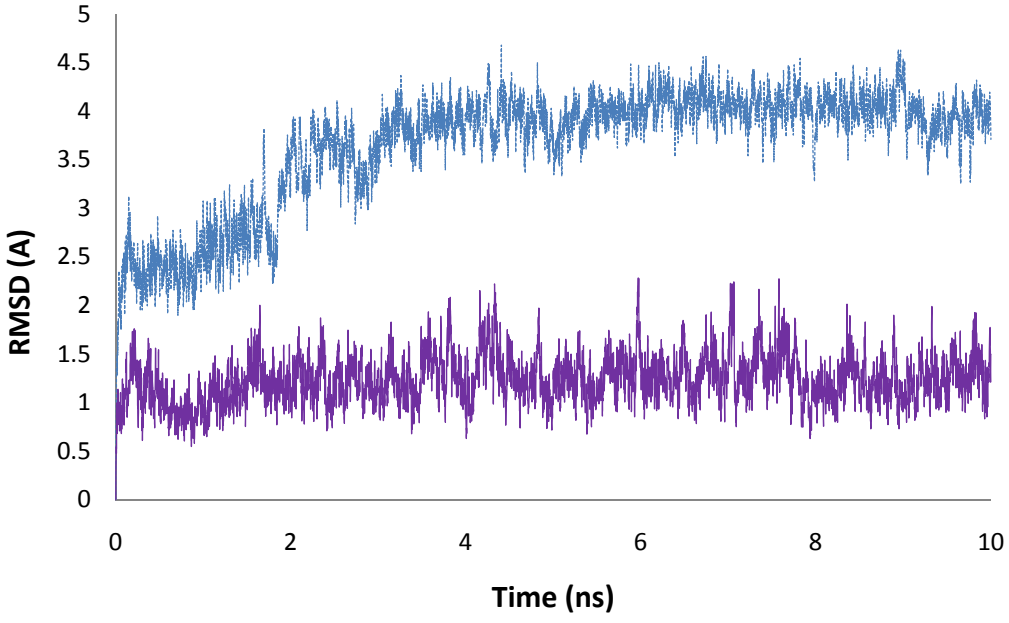



Figure 2.15: The central base pair RMSDs calculated over 10 ns of simulation for both the standard tetraloop (805 standard tetraloop in solid purple line) and the perfect d2 tetraloop (cutd2-tetraloop in dotted blue line).

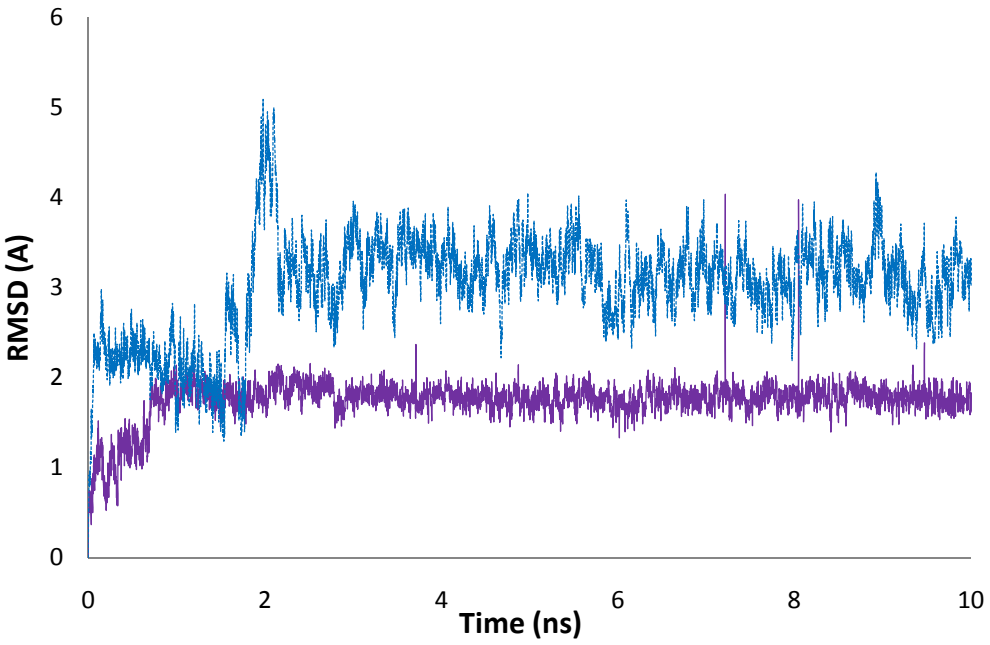



Figure 2.16: The closing base pair RMSDs calculated over 10 ns of simulation for both standard tetraloop (805 standard tetraloop in solid purple line) and the perfect d2 tetraloop (cutd2-tetraloop in dotted blue line).

The RMSDs of the stem base pairs showed that the terminus base pair of the perfect d2 tetraloop deviated the most (with a mean of 5.18 Å and a standard deviation of 1.07 Å) during the simulation compared to the standard tetraloop (with a mean of 1.55 Å and a standard deviation of 0.34 Å, Figure 2.14), followed by the central base pair (with a mean of 3.78 Å and a standard deviation of 0.61 Å for cutd2-tetraloop and a mean of 1.23 Å and a standard deviation of 0.25 Å for 805-standard tetraloop, Figure 2.15). The closing base pair of the perfect d2 tetraloop deviated the least (with a mean of 3.12 Å and a standard deviation of 0.56 Å, Figure 2.16), which is consistent with the hydrogen bonding interaction calculation conducted in the next section. However, compared to the standard tetraloop (the closing base pair RMSD of 805 standard tetraloop has a mean of 1.74 Å and a standard deviation of 0.22 Å), the closing base pair of the perfect d2 tetraloop deviated twice as much from its starting structure as the closing base pair of a standard tetraloop. All three stem base pairs statistics were

summarized in Table 3.2.

The RMSDs comparisons of the terminus base pair, the central base pair, and the closing base pair between cutd2-tetraloop and 805-standard tetraloop showed that the perfect d2 tetraloop had significantly higher RMSDs for all three base pairs in the stem, therefore validated again hypothesis I that the perfect d2 tetraloop is not a stable structure with no defects in its stem. The RMSDs of each base pair in the stem calculated over the simulation time also showed us that when the perfect d2 tetraloop became unstable during the simulation, the terminus base pair was the base pair that opened up first and deformed the most, followed by the central base pair, and then the closing base pair. The observation was consistent with the hypothesis that the loop forms first followed by the stem during tetraloop folding. This issue is further investigated in Chapter IV.

#### *2.3.3.2 Hydrogen Bond and Base Pairing Tracking*

Hydrogen bonding interaction is the molecular interaction that underlies the base pairing in both the loop and the stem of RNA molecules. Appropriate geometrical correspondence of hydrogen bond donors and acceptors allows only the "right" pairs to form. Base stacking is an even more important factor for nucleic acids stem stabilization. In this section we mainly focused our discussion on the conservation of hydrogen bonding interactions both in the loop and in the stem during the simulation for both the standard tetraloop and the perfect d2 tetraloop to compare their stability.

*2.3.3.2 a) Stem base pairing*

Here the distance between the proton donor and the acceptor of two complementary bases of each base pair in the stem was measured and plotted over the simulation time. We used 3.0 Å as the cutoff distance to form hydrogen bond. When the distance between the donor and acceptor is less than 3.0 Å a hydrogen bond is considered formed.

The distance of proton donor and acceptor of all three base pairs in the stem of 805 standard tetraloop were plotted in Figure 2.17, Figure 2.18 and Figure 2.19 to indicate the hydrogen bonding conservation in the stem during the simulation. The terminus base pair of 805-standard tetraloop, Guanine of residue 802 and Cytosine of residue 811 (Figure 2.6), formed three hydrogen bonds: 802G(H21)-811C(O2), 802G(O6)-811C(H41), and 802G(H1)-811C(N3). Figure 2.17 showed that the hydrogen bonding interactions in the terminus base pair were conserved very well throughout the simulation.

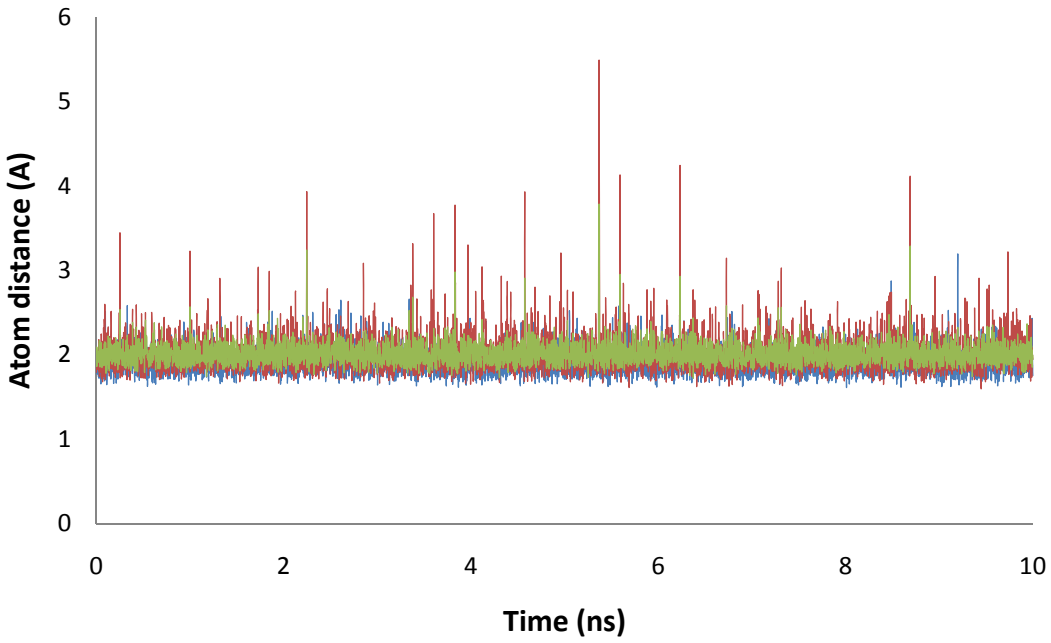



Figure 2.17: The terminus base pair of 805 standard tetraloop. Green line is the atom distance between H1 on 802 Guanine and N3 on 811 Cytosine. Red line is the atom distance between O6 on 802 Guanine and H41 on 811 Cytosine. Blue line is the atom distance between H21 on 802 Guanine and O2 on 811 Cytosine.

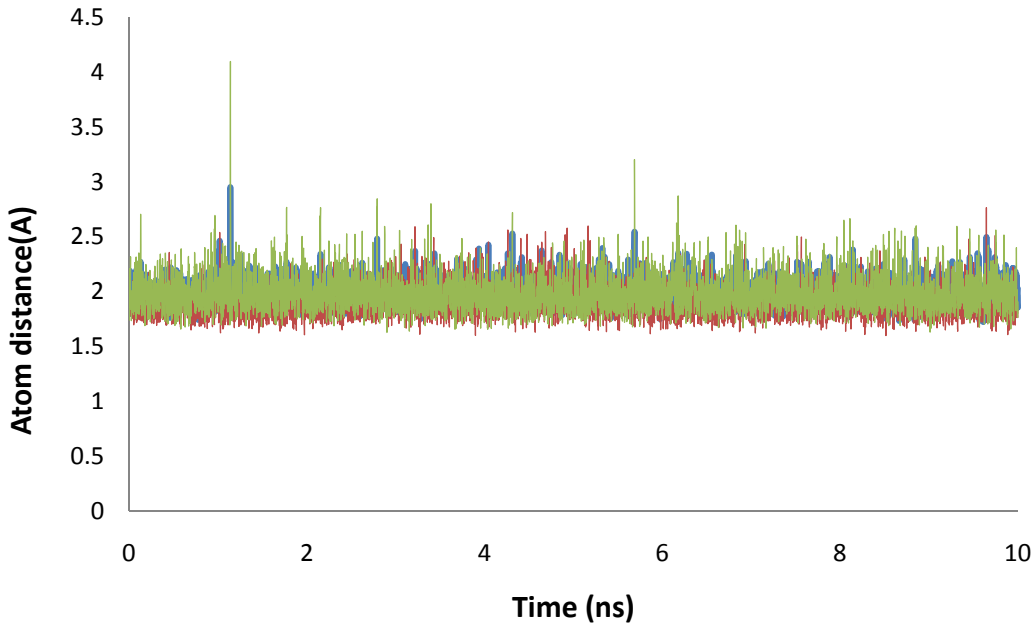



Figure 2.18: The central base pair of 805 standard tetraloop. Green line is the atom distance between H41 on 803 Cytosine and O6 on 810 Guanine. Red line is the atom distance between O2 on 803 Cytosine and H21 on 810 Guanine. Blue line is the atom distance between N3 on 803 Cytosine and H1 on 810 Guanine.

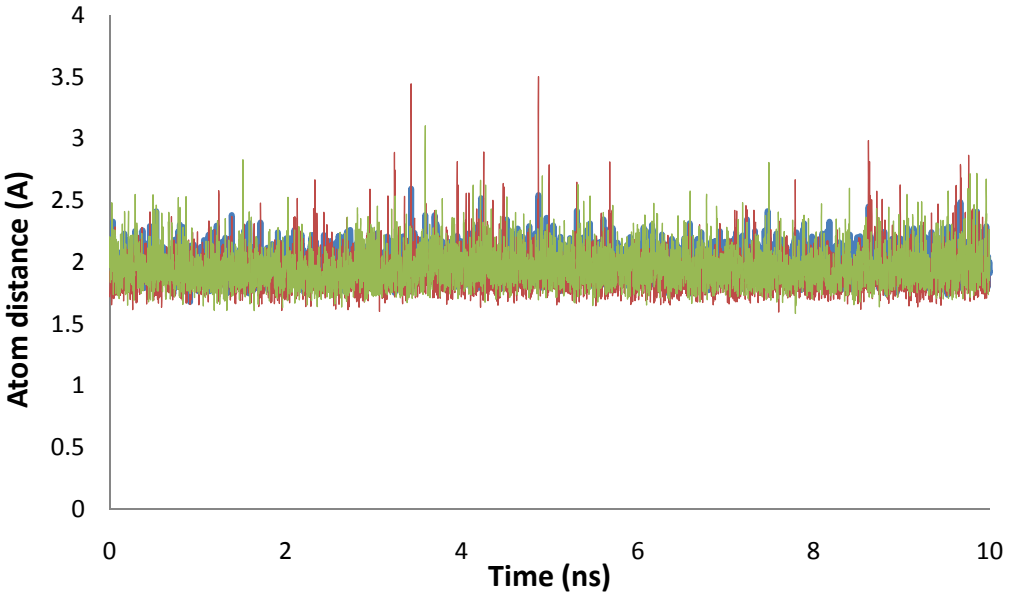



Figure 2.19: The closing base pair of 805 standard tetraloop. Green line is the atom distance between H41 on 804 Cytosine and O6 on 809 Guanine. Red line is the atom distance between O2 on 804 Cytosine and H21 on 809 Guanine. Blue line is the atom distance between N3 on 804 Cytosine and H1 on 809 Guanine.

The Central base pair of 805-standard tetraloop, Cytosine of residue 803 and Guanine of residue 810 (Figure 2.6), also formed three hydrogen bonds: 803C(H41)-810G(O6), 803C(H41)-810G(O6), and 803C(O2)-810G(H21). Figure 2.18 showed that the hydrogen bonding interactions in the terminus base pair of 805-standard tetraloop were also conserved very well throughout the simulation. The atom distances calculated for the closing base pair of the 805-standard tetraloop again showed well conserved hydrogen bonding interactions in the closing base pair (Figure 2.19). The cytosine of residue 804 and Guanine of residue 809 formed three hydrogen bonds (Figure 2.6): 804C(H41) – 809G(O6), 804C(H41) – 809G(O6), and 804C(O2) – 809G(H21). Hydrogen bonding statistics for each base pair in the stem of 805-standard tetraloop were summarized in Table 2.1.

Table 2.1: 805-standard tetraloop hydrogen bonding statistics for each base pair in the stem.

| | | Terminus BP | Central BP | Closing BP |
|---|---|---|---|---|
| **Hbond1** | Mean | 2.010037 | 1.982281 | 1.953681 |
| | Stdev | 0.121244 | 0.164377 | 0.151517 |
| **Hbond2** | Mean | 1.906903 | 1.984821 | 1.983278 |
| | Stdev | 0.138028 | 0.101902 | 0.097289 |
| **Hbond3** | Mean | 2.012249 | 1.893152 | 1.906756 |
| | Stdev | 0.219427 | 0.128908 | 0.151308 |

Figure 2.17 – 2.19 and Table 2.1 showed that all three bases in the stem of standard tetraloop were paired perfectly throughout 10 ns of MD simulation. The atom distances calculated for all three base pairs were significantly lower than 3 Å indicating very well conserved hydrogen bonding interactions for all three base pairs in the stem.The conservation of all three base pairing 802G-811C, 803C-810G, and 804C-809G in the stem of 805 standard tetraloop demonstrated the stability of a standard tetraloop during the simulation. It also provided a benchmark for the following stem base pairing analysis of the perfect d2 tetraloop (cutd2-tetraloop).

The distance between the proton donor and the acceptor of all three base pairs in the stem of the cutd2-tetraloop were calculated and plotted in Figure 2.20, Figure 2.21 and Figure 2.22. All hydrogen bonding statistics between each base pair in the stem were summarized in Table 2.2.

Figure 2.20 and Figure 2.21 showed that both the terminus base pairs and the central base pairs in the stem of cutd2-tetraloop started to open soon after the

simulation was started. The terminus base pair has a C – G pair (Cytosine of residue 311 and Guanine of residue 319 (Figure 2.5)) and should form three hydrogen bonds: 311C(N31)-319G(H1), 311C(H41)-319G(O6) and 311C(O2)-319G(H21). The central base pair should form two hydrogen bonds: 312U(O4)-318A(H61) and 312U(H3)-318A(N1).

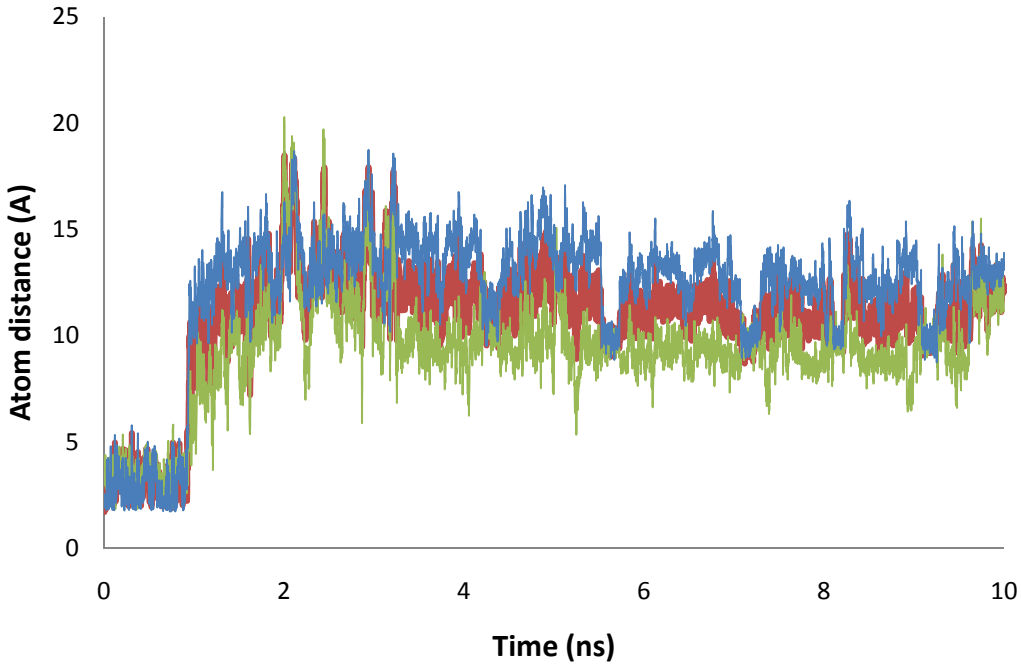



Figure 2.20: The terminus base pair of cutd2-tetraloop. Green line is the atom distance between O2 on 311 Cytosine and H21 on 319 Guanine. Red line is the atom distance between H41 on 311 Cytosine and O6 on 319 Guanine. Blue line is the atom distance between N3 on 311 Cytosine and H1 on 319 Guanine.

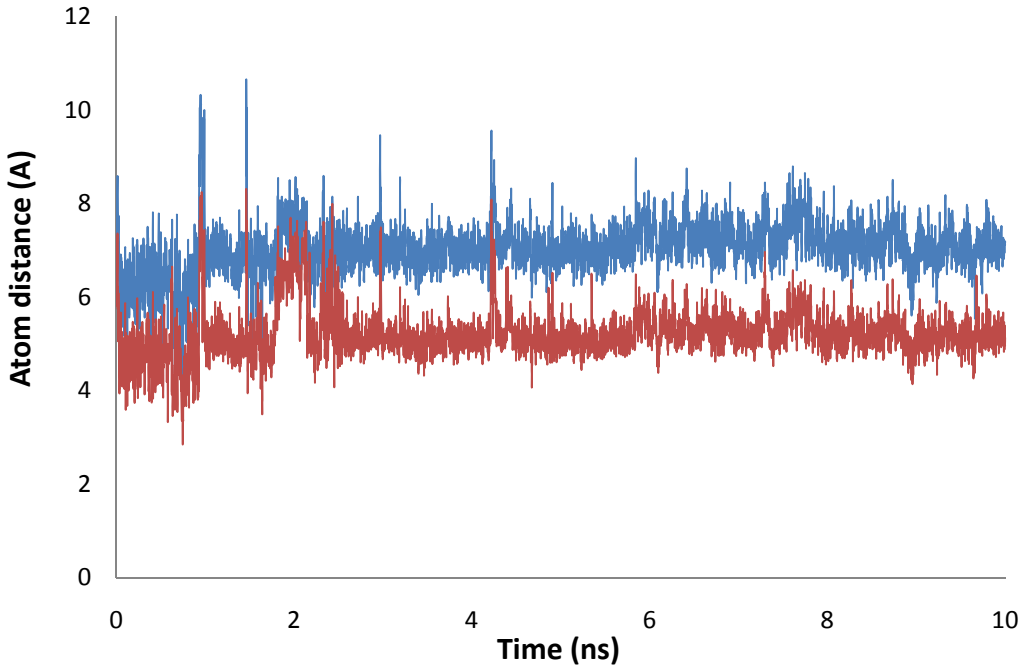



Figure 2.21: The central base pair of cutd2-tetraloop. Red line is the atom distance between H3 on 312 Uracil and N1 on 318 Adenine. Blue line is the atom distance between O4 on 312 Uracil and H61 on 318 Adenine.

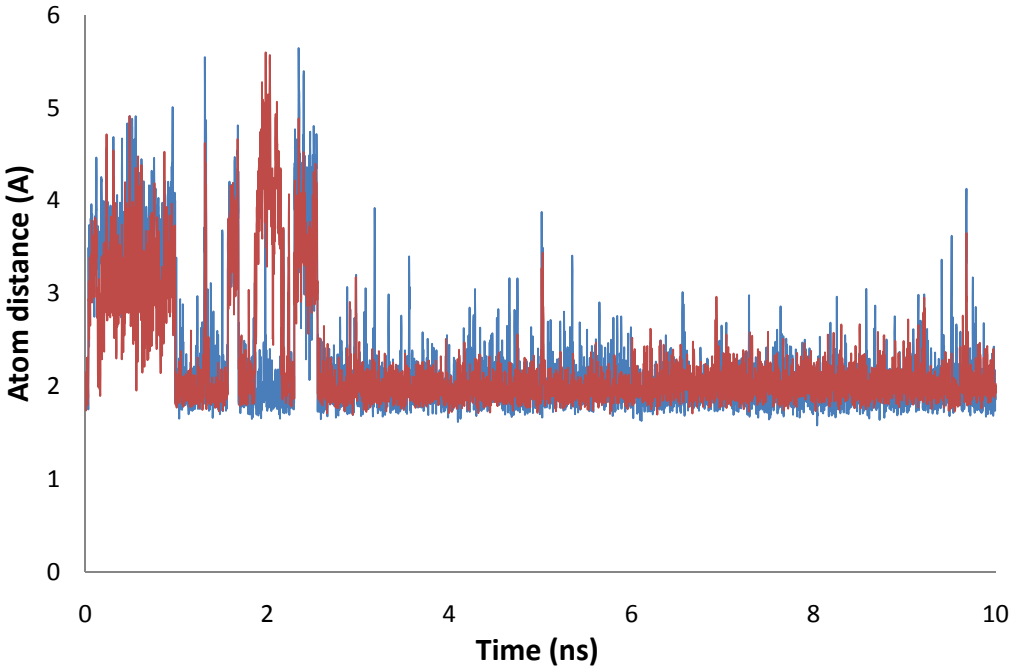



Figure 2.22: The closing base pair of 314-d2-tetraloop. Red line is the atom distance between H3 on 313 Uracil and N7 on 317 Adenine. Blue line is the atom distance between O4 on 313 Uracil and H61 on 317 Adenine.

The closing base pair also has a U – A pair and should form two hydrogen bonds: 313U(O4)-317A(H61) and 313U(H3)-317A(N7). Figure 2.22 showed that the closing base pair Uracil of residue 313 and Adenine of residue 317 opened up during the first 1 ns of simulation but somehow reformed those two hydrogen bonds during the second nanosecond of simulation. After 2 ns, the two closing bases remained pairing for the rest of the simulation.

The calculations of the hydrogen bonding interactions between each of the stem base pairs once again exhibited that the perfect d2 tetraloop, cutd2-tetraloop in this case, could not remain stable during the simulation, which again validated hypothesis I that a perfect d2 tetraloop with no defects in the stem is not stable relative to the standard tetraloop. Consistent with the RMSD calculations of the terminus base pair, the central base pair and the closing base pair, the results of hydrogen bonding interactions calculation for each stem base pair supports the theory that the loop forms first followed by stem during tetraloop formation.

Table 2.2: cutd2-tetraloop hydrogen bonding statistics for each base pair in the stem.

| | | Terminus BP | Central BP | Closing BP |
|---|---|---|---|---|
| **Hbond1** | Mean | 12.07537 | 5.216212 | 2.239202 |
| | Stdev | 3.290574 | 0.532055 | 0.596617 |
| **Hbond2** | Mean | 10.74436 | 7.045635 | 2.232802 |
| | Stdev | 2.817223 | 0.559375 | 0.640074 |
| **Hbond3** | Mean | 9.303866 | | |
| | Stdev | 2.573936 | | |

*2.3.3.2 b) Intra-loop base pairing*

Williams and coworkers [1] documented the consensus intra-loop hydrogen bonding interactions in the standard tetraloop. Their observations on hydrogen bonding interactions are consistent with expectations for 'GNRA' tetraloops [17] and U turns. They found that hydrogen bonding interaction of O2' of residue j-1 with cross-loop base atoms of residue j+1 are the most enduring one throughout the database. Twenty of 21 standard tetraloops form these hydrogen bonds in the database. Hydrogen bonding interactions also exist between j-1 and j+2. The base of residue j-1 can form hydrogen bonding with backbone atom O2P of residue j+2. Williams found that 14 of 14 standard tetraloops with G at j-1 show a hydrogen bond from either the N1 or the N2 of G at j-1 to the O2P of residue j+2, or both. The hydrogen bonding interactions between two bases of residue j-1 and residue j+2 were observed less frequently.

As a result, to compare the stability of the standard tetraloop and the perfect d2 tetraloop, the hydrogen bonding interactions of O2' of residue j-1 with the cross-loop base atoms of residue j+1 and the hydrogen bonding interactions between base of residue j-1 and backbone atom O2P of residue j+2 during the simulation were both calculated for both 805-standard tetraloop and cutd2-tetraloop.

Figure 2.23 shows the intra-loop hydrogen bonding interactions between residue j-1 and j+1 for both 805-standard tetraloop and cutd2-tetraloop. In the 805-standard tetraloop, the hydrogen bonding interaction between residue 805G at j-1 and residue 807A at j+1 were conserved very well throughout 10ns of simulation: the average distance between H2' of Guanine on residue 805 and N7 of Adenine on residue 807 is 1.98 Å (Figure 2.23). The molecular dynamics simulation on the 805-standard tetraloop generated results that are very consistent with the data mining results from the datebase and showed a very conserved hydrogen bonding interaction between residue

j-1 and residue j+1.

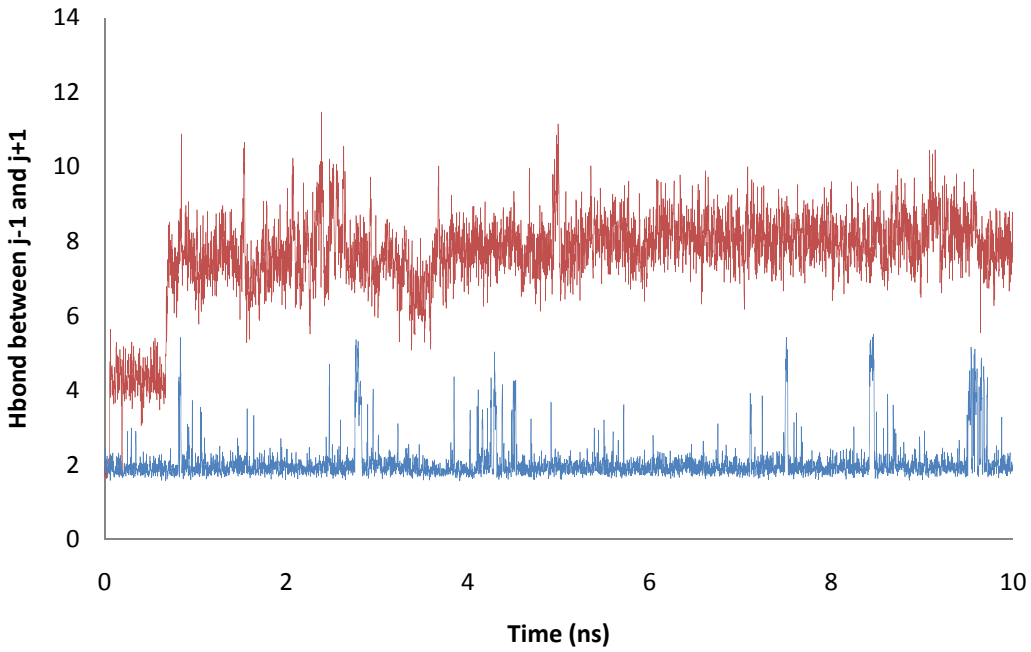



Figure 2.23: Intra-loop hydrogen bonding interaction (j-1 and j+1) comparison between 805-standard tetraloop (blue) and cutd2-tetraloop (red). The distance between H2' on Guanine of residue 805 at position j-1 and N7 on Adenine of residue 807 at position j+1 in blue. The distance between H2' on Guanine of residue 314 at position j-1 and N7 on Adenine of residue 316 at position j+1 in red.

In contrast, the calculated atom distance between H2' on Guanine of residue 314 at position j-1 and N7 on Adenine of residue 316 at position j+1 indicated the hydrogen bonding interaction between residue j-1 and residue j+1 in cutd2-tetraloop was not conserved during the simulation. Figure 2.24 showed the intra-loop hydrogen bonding interactions between residue j-1 and j+2 for both 805-standard tetraloop and cutd2-tetraloop.

In the 805-standard tetraloop, Guanine of residue 805 (position j-1) also formed hydrogen bonding interaction with the Adenine of residue 808 (position j+2)

throughout 10 ns of simulation. The average distance between H21 of the Guanine on residue 805 and O2P of the Adenine on residue 808 is 1.96 Å (Figure 2.24). Again the molecular dynamics simulation on 805-standard tetraloop generated results that are very consistent with the data mining results and showed a very conserved hydrogen bonding interaction between base of residue j-1 and backbone atom O2P of j+2.

In contrast, the calculated atom distance between H1 on Guanine of residue 314 at position j-1 and O2P on Adenine of residue 317 at position j+2 indicated once again that the hydrogen bonding interaction between residue j-1 and residue j+2 in cutd2-tetraloop was not conserved during the simulation.

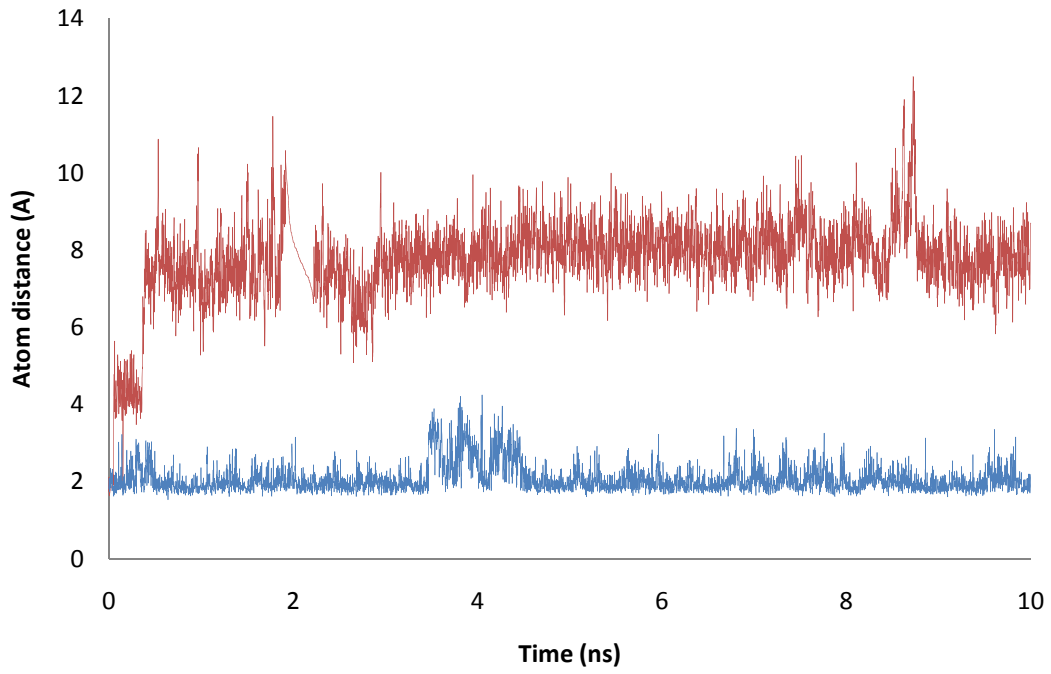



Figure 2.24: Intra-loop hydrogen bonding interaction (j-1 and j+2) comparison between 805-standard tetraloop (blue) and cutd2-tetraloop (red). The distance between H21 on Guanine of residue 805 at position j-1 and O2P on Adenine of residue 808 at position j+2 in blue, the distance between H1 on Guanine of residue 314 at position j-1 and O2P on Adenine of residue 317 at position j+2 in red.

In summary, calculations of intra-loop hydrogen bonding interactions (the one between j-1 and j+1 and the one between j-1 and j+2) from 805-standard tetraloop simulation showed consistency with data mining results. Comparisons of these intra-loop hydrogen bonding interactions between 805-standard tetraloop and cutd2-tetraloop again showed that the perfect d2 tetraloop could not remain stable during the simulation and validated hypothesis I.

*2.3.3.3 Torsion Angle Analysis*

Torsion angle analysis was conducted for stability comparison. Each RNA residue has six backbone torsion angles (α, β, γ, δ, ε, δ), and a glycosidic torsion angle (χ) and ribose pseudorotation angle (P) (Figure 2.25). The angle α of residue i is the $O3'_{i-1}$-$P_i$-$O5'_i$-$C5'_i$ torsion angle, β is the $P_i$-$O5'_i$-$C5'_i$-$C4'_i$ torsion angle, etc. as generally defined for nucleic acids [18]. The conformation of an *n*-residue RNA molecule can be specified by *n* sets of (α, β, γ, δ, ε, δ, χ, and P). For computational efficiency the usual torsional format of -180° to +180° was converted to 0° to 360°.

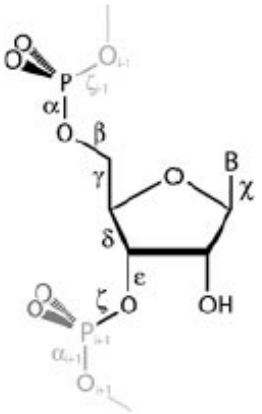


Figure 2.25: Ribonucleotide torsion angles used to specified conformation [18].

*2.3.3.3 a) Torsion Angle and Binning Method*

Here we utilize the binning method proposed by Williams and coworkers [19]. In their method, each torsion angle of a given residue was empirically 'binned' by allocating it to the appropriate envelope, which was then assigned a discrete integer value. Thus, the continuous torsion angle data was converted into integers, which specify the correspondence of torsion angle to a Gaussian envelope.

Williams and coworkers found that β, ε, and χ do not contribute information to the conformational description and therefore were ignored. In addition, δ and P are, by geometric definition [20, 21], correlated and therefore were eliminated to avoid redundancy. As a result, only four torsion angles (α, γ, δ, and ζ), which they called four 'conformational identifier' angles were considered. Each residue was assigned a sequence of four integers $n_\alpha$, $n_\gamma$, $n_\delta$, and $n_\zeta$, where each digit denotes the Gaussian envelope to which a torsion angle belongs. The range of each envelope for the four identifier angles is given in Table 2.3 [19].

Table 2.3: Bin assignments and torsion ranges [19].

| | Bin 1 | Bin 2 | Bin 3 | Bin 4 |
|---|---|---|---|---|
| α | 40–90 | 135–190 | 260–330 | other |
| γ | 35–75 | 150–200 | 260–320 | other |
| δ | 68–93 | 130–165 | other | |
| ζ | 255–325 | other | | |

These definitions lead to 4x4x3x2=96 possible conformational states. However only 37 of these states are populated in the HM LSU 23S rRNA (bins occupied by more than five residues were considered to be populated in their work). ASCII symbols with corresponding bin numbers and observation frequencies were

summarized by Williams and coworkers in Table 2.4.

Table 2.4: ASCII symbols, bin numbers and observation frequencies [19].

| Ascii letter[a] | Bin number | Frequency |
|---|---|---|
| a | 3111 | 1709 |
| e | 3112 | 169 |
| r | 3122 | 124 |
| i | 2211 | 103 |
| o | 2111 | 58 |
| t | 4111 | 48 |
| n | 1111 | 37 |
| s | 2122 | 34 |
| l | 1211 | 31 |
| c | 3121 | 30 |
| u | 4211 | 28 |
| d | 1121 | 26 |
| p | 4122 | 21 |
| m | 1122 | 21 |
| h | 3411 | 18 |
| g | 1322 | 18 |
| b | 1112 | 14 |
| f | 3211 | 14 |
| y | 4112 | 13 |
| w | 2212 | 11 |
| k | 4121 | 11 |
| v | 3212 | 10 |
| x | 3222 | 10 |
| z | 1331 | 9 |
| j | 4222 | 9 |
| q | 3321 | 8 |
| 1 | 1212 | 8 |
| 2 | 3422 | 8 |
| 3 | 4311 | 8 |
| 4 | 4411 | 8 |
| 5 | 2121 | 7 |
| 6 | 3322 | 7 |
| 7 | 2222 | 7 |
| 8 | 2411 | 7 |
| 9 | 1311 | 7 |
| 0 | 1221 | 7 |
| + | 3311 | 6 |

The binning of the ASCII code for a four-residue tetraloop motif is ‘aoaa’, which represents bin number 3111, 2111, 3111, and 3111 respectively. This pattern was observed 24 times in the HM 23S rRNA (Table 2.5). The standard tetraloop 805 s-TL and d2 tetraloop with insertion (314-d2-tetraloop) studied are also in the table. Therefore simulation results on backbone torsion angle of standard tetraloop and

perfect d2-tetraloop can be compared with data mining results from the database.

Table 2.5: Tetraloop identified by the binning method [19].

| | Starting residue | Sequence[a] | Binning ASCII code |
|---|---|---|---|
| 1 | 252 | CUCAC | aaoa |
| 2 | 313 | UGGAA | aaoa |
| 3 | 468 | UGUGA | aaoa |
| 4 | 505 | CGAAA | aaoa |
| 5 | 624 | UUUGA | aaoa |
| 6 | 690 | GGAAA | aaoa |
| 7 | 804 | CGAAA | aaoa |
| 8 | 1054 | GGUAA | aaoa |
| 9 | 1197 | GUAAC | aaoa |
| 10 | 1326 | UGAAA | aaoa |
| 11 | 1388 | UGAGA | aaoa |
| 12 | 1468 | GCAAC | aaoa |
| 13 | 1499 | UUAAU | aaoa |
| 14 | 1595 | GUAAU | aaoa |
| 15 | 1628 | GGAAA | aaoa |
| 16 | 1706 | GGCGA | aaoe[b] |
| 17 | 1748 | UUCGG | aaoa |
| 18 | 1793 | CGGAA | aaoa |
| 19 | 1808 | CGCAG | aaoa |
| 20 | 1862 | CGCAA | aaoa |
| 21 | 1991 | AUCAG | aaoa |
| 22 | 2248 | CGGGA | aaoa |
| 23 | 2411 | CGAAA | aaoa |
| 24 | 2629 | CGUGG | aaoa |
| 25 | 2695 | CGAGA | aaoa |

*2.3.3.3 b) Torsion Angle Analysis of 805-Standard Tetraloop and Cutd2-Tetraloop*

Ten ns of Molecular Dynamics simulation of 805-standard tetraloop showed a very stable structure for the standard tetraloop. Four backbone torsion angles (α, γ, δ, and ζ) of four residues in the loop were measured throughout the simulation with their mean and standard deviation recorded in Table 2.6. For each torsion angle the first row is the starting torsion angle, the second row is the mean of the torsion angle during the simulation, and the third row is the standard deviation. The four columns represent torsion angles of four residues j-1, j, j+1, and j+2 in the loop region.

The binning of the ASCII code for the 805 standard tetraloop is ‘aoaa’, which represents bin number 3111, 2111, 3111, and 3111 respectively. According to bin

assignments and torsion angles (Table 2.3), 3111 means α with range of 260° – 330°, γ with range of 35° – 75°, δ with range of 68° – 93°, and ζ with range of 255° – 325°. And 2111 means α with range of 135° – 190°, γ with range of 35° – 75°, δ with range of 68° – 93°, and ζ with range of 255° – 325°.

Table 2.6: Four backbone torsion angle for four residues in the loop of 805-standard tetraloop.

| | | 805G | 806A | 807A | 808A |
|---|---|---|---|---|---|
| α | Starting Point | 290.05 | 185.09 | 298.06 | 310.36 |
| | Mean | 292.09 | 86.44 | 294.07 | 285.66 |
| | Stdev | 10.98 | 41.20 | 18.77 | 11.12 |
| γ | Starting Point | 60.96 | 43.89 | 52.616 | 40.57 |
| | Mean | 56.10 | 175.21 | 65.05 | 59.57 |
| | Stdev | 9.74 | 39.30 | 10.03 | 9.31 |
| δ | Starting Point | 69.98 | 81.24 | 77.80 | 92.49 |
| | Mean | 73.79 | 76.22 | 80.61 | 80.90 |
| | Stdev | 7.64 | 9.40 | 6.70 | 8.06 |
| ζ | Starting Point | 286.13 | 291.57 | 307.66 | 296.92 |
| | Mean | 270.78 | 286.29 | 296.44 | 240.57 |
| | Stdev | 13.94 | 64.64 | 10.05 | 12.84 |

Backbone torsion angle analysis of the 805-standard tetraloop from the simulation showed again that the standard tetraloop is a very stable structure. Most of the backbone torsion angles (805G, 807A and 808A) calculations were consistent with the empirical observations. Only backbone torsion angles α and γ of residue 806A were not consistent with empirical observation.

In contrast, 10 ns of MD simulation of cutd2-tetraloop showed a much less stable structure of the perfect d2 tetraloop with no defects in the stem. Four backbone torsion angles (α, γ, δ, and ζ) of four residues in the loop were also measured

throughout the simulation with their mean and standard deviation recorded in Table 2.7. Again, for each torsion angle the first row is starting torsion angle, the second row is the average torsion angle during the simulation, and the third row is the standard deviation. Because there are only three residues in the loop region of a perfect d2 tetraloop, the three columns represent torsion angles of three residues j-1, j, and j+1.

The binning of the ASCII code for 314-d2-tetraloop is 'aoa', which represents bin number 3111, 2111, and 3111 respectively. Therefore the same binning of ASCII code was used for cutd2-tetraloop as well. According to bin assignments and torsion angles (Table 2.3), 3111 means α with range of 260° – 330°, γ with range of 35° – 75°, δ with range of 68° – 93°, and ζ with range of 255° – 325°. And 2111 means α with range of 135° – 190°, γ with range of 35° – 75°, δ with range of 68° – 93°, and ζ with range of 255° – 325°.

Table 2.7: Four backbone torsion angle for three residues in the loop of cutd2-tetraloop.

| | | 314G | 315G | 316A |
|---|---|---|---|---|
| α | Starting Point | 252.55 | 89.432 | 299.21 |
| | Mean | 227.212 | 139.99 | 163.50 |
| | Stdev | 77.14 | 81.14 | 100.10 |
| γ | Starting Point | 66.65 | 91.44 | 42.80 |
| | Mean | 111.61 | 173.08 | 134.27 |
| | Stdev | 69.00 | 79.03 | 81.85 |
| δ | Starting Point | 78.923 | 114.42 | 86.40 |
| | Mean | 76.66 | 139.78 | 124.68 |
| | Stdev | 8.16 | 18.74 | 21.65 |
| ζ | Starting Point | 317.08 | 246.65 | 310.76 |
| | Mean | 279.35 | 222.55 | 284.83 |
| | Stdev | 12.74 | 22.45 | 15.22 |

Backbone torsion angle analysis of cutd2-tetraloop from the MD simulation showed again that a perfect d2 tetraloop with no stem defects did not remain stable. The average value of most of the torsion angles fell outside of the range with very large standard deviations as well. When comparing the torsion angle analysis of cutd2-tetraloop to that of 805 standard tetraloop we can conclude perfect d2 tetraloop is not a stable structure.

## *2.4 Conclusions*

MD simulations of the 805-standard tetraloop generated results, such as intra-loop hydrogen bonding and loop residue backbone torsion angle distribution, that were consistent with data mining results. Consequently the simulation parameters and algorithms used were proven to generate accurate results and will continue to be used in the following chapters.

MD simulations of cutd2-tetraloop validated hypothesis I that the perfect d2 tetraloop with no defects in the stem is not a stable structure. By the end of 10 ns of MD simulation the structure of cutd2-tetraloop unraveled opening the stem and deforming the loop. RMSDs of all residues, the loop region and the stem region of the cutd2-tetraloop were significantly higher compared to those of 805-standard tetraloop. Hydrogen bonding interactions both in the loop and in the stem were not conserved well during the simulation, with the exception of the closing base pair. This suggests the closing base pair is relatively stable. Torsion angle analysis of cutd2-tetraloop

supported the conclusion that the perfect d2 tetraloop is not stable. Further investigation on the stabilizing effect of the stem defects is conducted in the next chapter.

## *2.5 References*


1. Hsiao, C., Mohan, S., Hershkovitz, E., Tannenbaum, A., & Williams, L. D. (2006). Single nucleotide RNA choreography. *Nucl. Acids Res*. 34 (5), 1481–1491.
2. Antao, V. P., Lai, S. Y. & Tinoco, I., Jr (1991). A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucl. Acids Res.* **19**, 5901–5905.
3. Antao, V. P. & Tinoco, I., Jr (1992). Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucl. Acids Res.* **20**, 819–824.
4. Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stomo, G., Daubentoncarafa, Y., Uhlenbeck, O. C., Tinoco, I., Jr, Brody, E. N. & Gold, L. (1988). CUUCGG hairpins–extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl Acad. Sci. USA,* **85**, 1364–1368.
5. Klein,D.J., Schmeing,T.M., Moore,P.B. and Steitz,T.A. (2001) The kink-turn: a new RNA secondary structure motif. EMBO J., 20, 4214–4221
6. Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 s resolution. Science, 289, 905–920
7. Lee,J.C., Cannone,J.J. and Gutell,R.R. (2003) The lonepair triloop: a new motif in RNA structure. J. Mol. Biol., 325, 65–83.
8. Cornell, W. D., Cieplak P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., & Kollman, P. A. A second generation

force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117 (19): 5179-5197, 1995.

9. Jorgensen, W.L., et al., Comparison of simple potential functions for simulating liquid water. Journal of Chemical Physics, 1983. 79: p. 926-935.
10. Humphrey, W., A. Dalke, and K. Schulten, VMD: visual molecular dynamics. J. Mol Graph, 1996. 14(1): p. 33-8, 27-8
11. Darden., T., D. York, and L. Pedersen, Paticle Mesh Ewald--an N.Log(N) method for Ewald sums in large systems. *Journal of Chemical Physics*, 1993. 98: p. 10089-10092.
12. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. *J. Chem. Phys.* 1995. 103: 8577-8593.
13. Ryckaert, J.P., G. Ciccotti, and H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constrains: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 1977. 23: p. 327-341.
14. *Berendsen, H. J. C.; Postma, JPM; Van Gunsteren, WF; Dinola, A; Haak, JR (1984). Molecular-Dynamics with Coupling to an External Bath. Journal of Chemical Physics 81 (8): 3684–3690.*
15. A. T. Br□unger. X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR. The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, 1992.
16. Harvey, S. C., Tan, R. K. Z., and Cheatham, T. E. (1998). The Flying Ice Cube: Velocity Rescaling in Molecular Dynamics Leads to Violation of Energy Equipartition. Journal of Computational Chemistry, 19 (7), 726-740.
17. Jucker,F.M., Heus,H.A., Yip,P.F., Moors,E.H. and Pardi,A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.*, 264, 968–980.
18. Saenger, W. (1984). *Principle of Nucleic Acid Structure*. Springer Verlag, New York, NY.
19. Hershkovitz, E., Tannenbaum, E., Howerton, S. B., Sheth, A., Tannenbaum, A. and Williams, L. D. (2003). Automated identification of RNA conformational motifs:

theory and application to the HM LSU 23S rRNA. *Nucleic Acids Research*, 31(21), 6249-6257.

20. Olson,W.K. (1982) Theoretical studies of nucleic acid conformation: Potential energies, chain statistics and model building. In Neidle,S. (ed.), *Topics in Nucleic Acid Structure*, Part 2. Macmillan Press Inc., London, UK, 1-79.
21. Sundaralingam,M. (1973) Conformation of biological molecules and polymers. In Bergmann,E.D. and Pullman,B. (eds), *The Jerusalem Symposium on Quantitative Chemistry and Biochemistry.* Academic Press, New York, NY, Vol. 5, 417-456.

# CHAPTER III

# STABILIZING EFFECTS OF STEM DEFECTS

## IN D2-TRATALOOPS

This chapter analyzes stabilizing effect of defects in the stem of the d2 tetraloop through Molecular Dynamics simulation of a sample structure of the d2 tetraloop with stem insertion residue, 314-d2-tetraloop.

## *3.1 Introduction*

### 3.1.1. Background

As discussed earlier in Chapter II, d2 tetraloops belong to the deletion family of DevLS [1] and are a tetraloop with residue j+2 of the standard tetraloop absent (Figure 2.3). The deletion of residue j+2 in the loop area does not appreciably change the atomistic positions of the remaining backbone of these tetraloops. However, the deletion at position j+2 was found to be correlated with adjacent helical distortions. In chapter II Molecular Dynamics simulations of cutd2-tetraloop showed that the perfect d2 tetraloop is not a stable structure without stem defects.

Our test database is still HM LSU 23S rRNA, the crystal structure of which has been determined to high resolution [2, 3]. D2 tetraloops abundantly exist not only in

this database but also in other databases, referred to as Lonepair Triloop [4]. This high abundance of d2 tetraloops raised the questions of *which factors stabilize this seemingly strained structure*.

### 3.1.2 Motivation

Gutell and coworkers [4] have also identified numerous examples of similar structure to the d2 tetraloop, referred to as Lonepair Triloop (LPTL). They found approximately half of the LPTL motifs are in close association with proteins. The majority of these LPTLs are positioned at sites in rRNAs that are conserved in three phylogenetic domains; a few of these occur in regions of rRNA associated with ribosomal function, including the presumed site of peptidyl transferase activity in the 23S rRNA.

As discussed in Chapter II, the d2 tetraloop has only three residues in the loop region and therefore has less hydrogen bonding and stacking interactions in the loop compared to the standard tetraloop. The stem of the d2-tetraloop does not form perfect base pairing and stacking like those in the stem of a standard tetraloop. Such a seemingly less stable structure has such an abundant existence and is associated with a lot of critical biological functions. Therefore it is very important to understand what makes this structure stable.

Williams and coworkers [1] have found deletion at the j+2 position is correlated with adjacent helical distortion. Unlike standard tetraloops, which have a healthy stem with perfect base pairing and stacking, d2-tetraloops always appear with some imperfections in the stem. In comparison to standard tetraloops, due to the missing residue j+2, d2-tetraloops form an unhealthy stem with adjacent helical imperfection such as insertion at position 3 (314, 625, 1387, 1992), clipping at position 2 (1187, 1809, 2598), base pairing disruption in the stem (1500, 1596) and unhinging

(1749) (Figure 2.4).

We propose the theory that this type of stem distortion or conformational perturbation is the reason for the abundant existence of the d2-tetraloop. The standard tetraloop is known to always form a healthy stem with nice base pairing and stacking in the stem. When a residue at position j+2 is missing, the whole structure is thought to be highly strained and therefore unstable. We take one example of d2-tetraloops with insertions as the stem distortion for instance. In a d2-tetraloop with insertions, by adding one or several insertion residues, the stem will be 'distorted' but somehow relaxed. This relaxation of stem compensates for the high energy from the highly strained loop area due to one missing residue. Therefore the insertion on the stem is believed to play an important role in stabilizing the d2-tetraloop structure by relieving some of the energy restrain in the loop area.

Here Molecular Dynamics simulations were performed on the d2 tetraloop with defects in the stem. The simulation results were compared to those of the perfect d2 tetraloop with no stem defects to reveal the stabilizing factor of d2 tetraloop and explain the high abundance of d2 tetraloop in the database.

### 3.1.3 Objectives

Determine the stabilizing factor of d2-tetraloop by running simulations on a sample structure of d2 tetraloops with some stem defects and comparing the simulation results to the perfect d2 tetraloop.

*Hypothesis II. Stem defects stabilize the d2-tetraloop*

In order to test Hypothesis II, we performed MD simulations on the d2 tetraloop with certain stem defects and compared it to the perfect d2 tetraloop in Chapter II. Simulations in Chapter II showed that the stability of the d2 tetraloop is related to stem defects due to the fact that the perfect d2 tetraloop without stem defects could not remain stable. Molecular Dynamics simulation of the d2 tetraloop with stem defects should also give more insights on how exactly the defected stem stabilizes the structure of d2 tetraloops so that the abundance of d2 tetraloops could be explained.

## *3.2 Methods*

### 3.2.1 Molecular Dynamics Simulation on 314-d2-tetraloop

The same tetraloop used in Chapter II to build cutd2-tetraloop, 314-d2-tetraloop from the database, was used as a sample structure to test our hypothesis. Again 314-d2-tetraloop is a d2-tetraloop with a single residue insertion in the stem. The inserted residue is between residue j+3 and j+4 on the 3' side of the stem. The number (314) represents the nucleotide number in HM 23S rRNA secondary structure and is also the first residue number (j-1) in the loop.

314-d2-tetraloop has a sequence of CUU(GGA)A(C)AG. It is an example of the d2-tetraloop with one residue insertion (318C) in the stem (Figure 3.1). Molecular dynamics simulation was performed on this structure to investigate the stabilizing effect of stem defects, the insertion residue in this case.

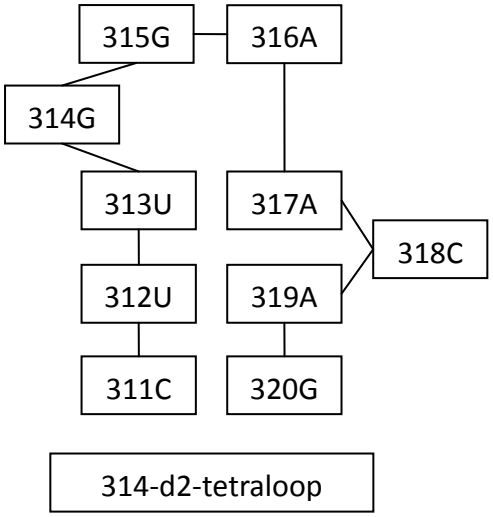



Figure 3.1. Secondary structure of 314-d2-tetraloop.

### 3.2.2 Simulation Protocol

Ten ns of NPT MD simulation in explicit water (TIP3P [5]) was performed following the same simulation protocol in Chapter II. Again AMBER94 [6] was used as the force field used.

## *3.3 Results and Discussions*

### 3.3.1 Simulation Results Comparison between the Perfect d2 Tetraloop and the d2 Tetraloop.

*3.3.1.1 Simulation Observations*

*3.3.1.1 a) Simulation Observations of 314-d2-tetraloop*

The molecular dynamics simulation showed that the d2 tetraloop with stem defects, 314-d2-tetraloop with insertion residue 318C in this case, remained very stable. Throughout 10 ns of MD simulation on 314-d2-tetraloop, the positions of backbone atoms did not change appreciably. The snapshot of the starting structure from the simulation looks almost the same as the snapshot of the end structure from the simulation (Figure 3.2).

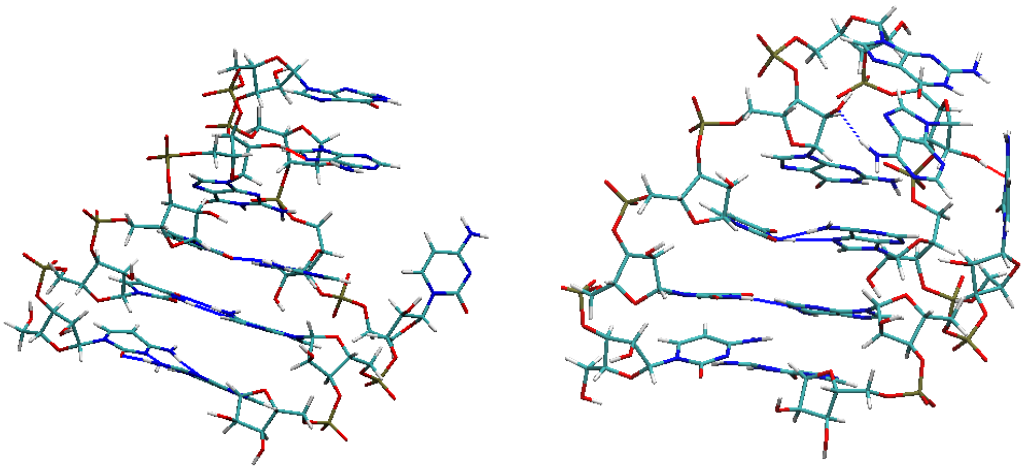


Figure 3.2: The snapshots of the start structure (left) and the end structure (right) from 10ns MD simulation of 314-d2-tetraloop.

In the loop area residue Guanine (314G) at position j-1, Guanine (315G) at position j, and Adenine (316A) at position j+1 stayed at their starting atomistic point throughout the simulation. RMSDs were calculated in the following section. Both

hydrogen bond interactions between residue 314G at position j-1 and residue 316A at position j+1, and 314G at j-1 and backbone atom on residue 317A at j+2, conserved very well throughout the simulation. Terminus base pair 311C-320G, central base pair 312U-319A, and closing base pair 313U-317A were pairing in the stem very well. Detailed hydrogen bonding interactions both in the loop and in the stem were calculated in the following section. It seemed bases also perfectly stacked on each side of the stem. There was no sign of stem unraveling or loop loosening. It seems that this 314-d2-tetraloop with an insertion residue 318C in the stem remained relatively stable throughout the simulation.

*3.3.1.1 b) Observation comparison between 314-d2-tetraloop and cutd2-tetraloop.*

Simulation of cutd2-tetraloop in Chapter II showed that the perfect d2 tetraloop is a very unstable structure compared to the standard tetraloop.

By comparing the ending structure of 314-d2-tetraloop and that of cutd2-tetraloop from the simulation (Figure 3.3), we can see that the loop structure and the stem structure of 314-d2-tetraloop remained intact with perfect base pairing and stacking throughout the simulation, while the structure of cutd2-tetraloop started to fall apart with opening stem and deformed loop. At this point we conclude qualitatively that the insertion residue in the stem played a critical role in stabilizing 314-d2-tetraloop, which validated our hypothesis II that stem defects stabilize d2 tetraloops.

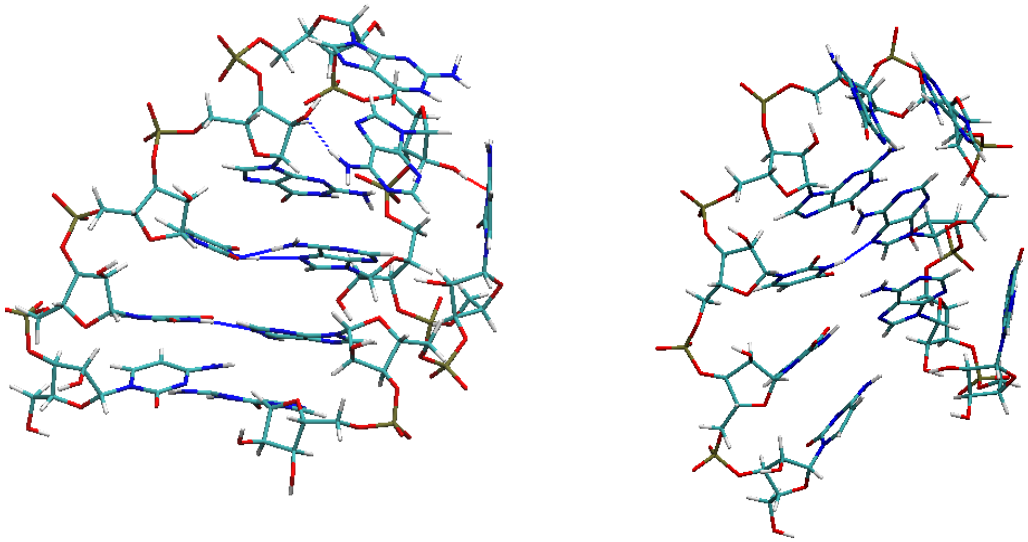


Figure 3.3: The ending structures of the 314-d2-tetraloop (left) and the cutd2-tetraloop (right) from the simulation.

*3.3.1.2 RMSD comparisons*

The root-mean-squared atomic deviations (RMSDs) of all residues, loop region and stem region of 314-d2-tetraloop were calculated and plotted on the same graph with RMSDs of cutd2-tetraloop to indicate the structural changes of both 314-d2-tetraloop and cutd2-tetraloop during the simulations (Figure 3.4, Figure 3.5 and Figure 3.6). The trajectories were superposed on the starting structure before the RMSDs were calculated.

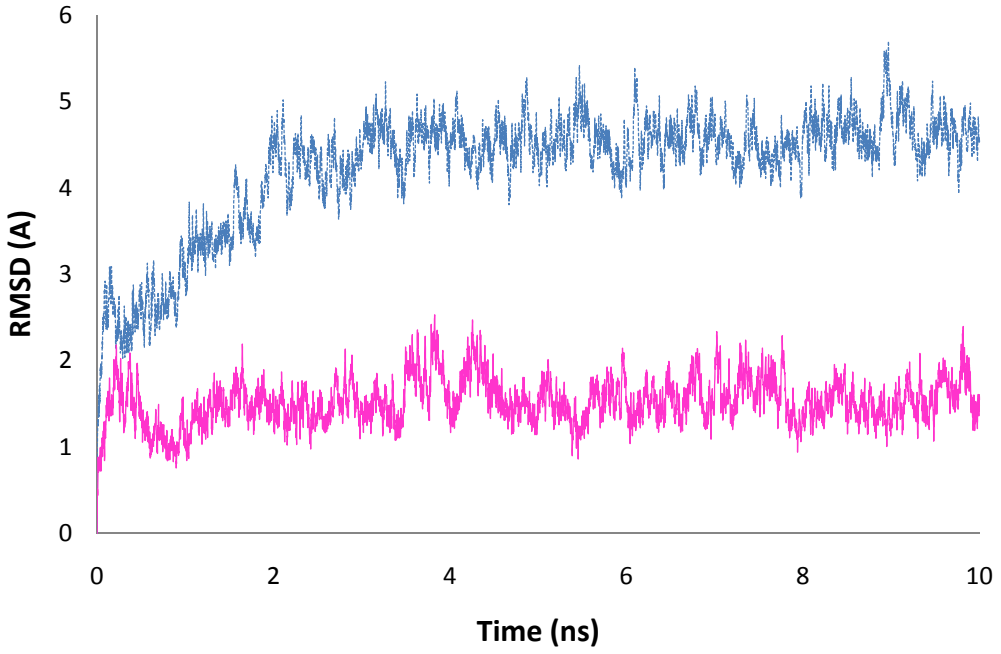



Figure 3.4: All residues RMSD calculation for 314-d2-tetraloop (pink solid line) and cutd2-tetraloop (blue dotted line).

All residues RMSD values (Figure 3.4) showed that the structure of 314-d2-tetraloop did not change considerably throughout the simulation, with a mean of 1.53 Å and a standard deviation of 0.27 Å. In contrast, RMSD values of cutd2-tetraloop were significantly higher with a mean and a standard deviation of 4.24 Å and 0.71 Å, indicating that the structure of cutd2-tetraloop deviated a lot from its starting point compared to 314-d2-tetraloop. The all residue RMSDs calculation comparison between 314-d2-tetraloop and cutd2-tetraloop implied that the perfect d2 tetraloop with no defects in the stem, cutd2-tetraloop with insertion residue 318C removed from the stem in this case, is not a stable structure compared to the d2 tetraloop with stem defects.

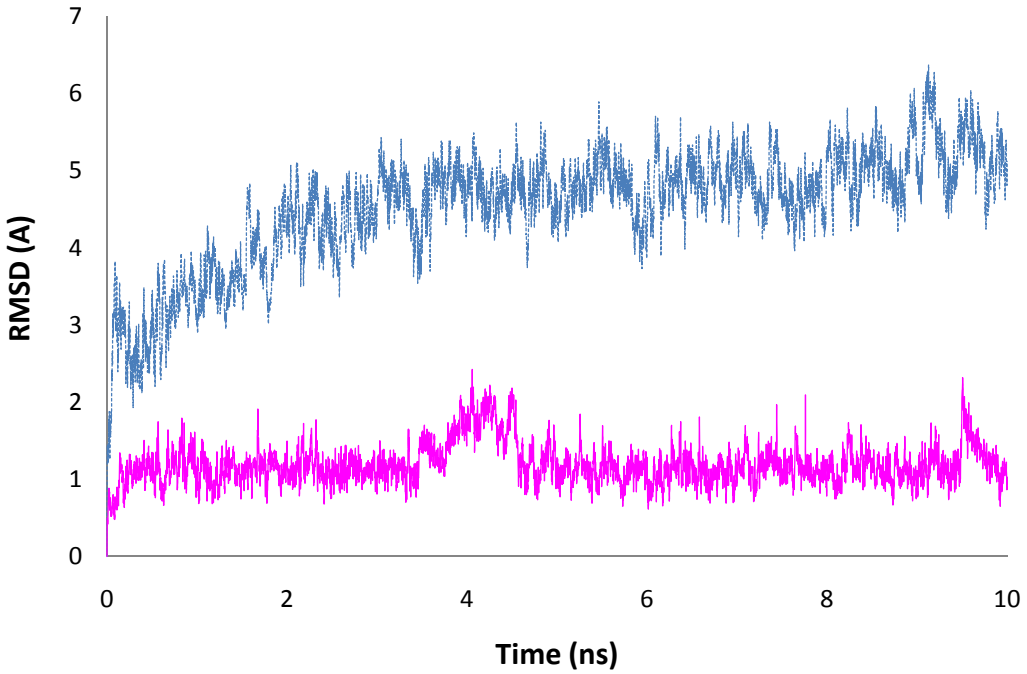



Figure 3.5: Loop RMSD calculation for 314-d2-tetraloop (pink solid line) and cutd2-tetraloop (blue dotted line).

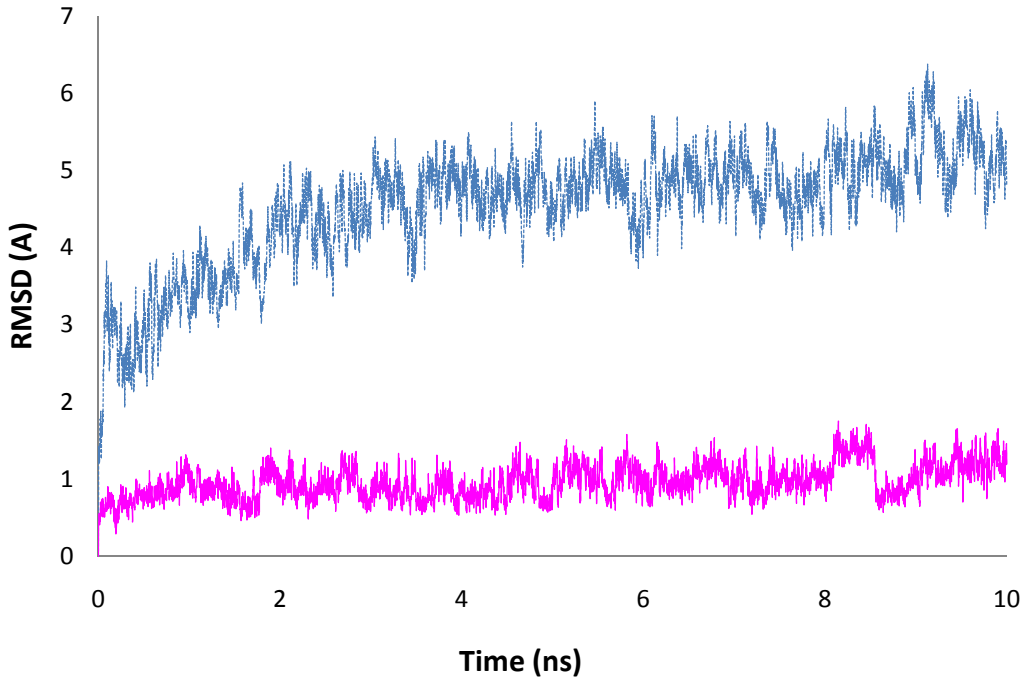



Figure 3.6: Stem RMSD calculation for 314-d2-tetraloop (pink solid line) and cutd2-tetraloop (blue dotted line).

Figure 3.5 and Figure 3.6 showed the RMSD comparison of loop and stem between 314-d2-tetraloop and cutd2-tetraloop. Again cutd2-tetraloop has significantly higher RMSDs for both loop and stem compared to 314-d2-tetraloop. This comparison once again suggested that the d2 tetraloop became unstable when the defects were removed from its stem.

Table 3.1 summarized the all residue, loop and stem RMSD statistics for 805-standard tetraloop, 314-d2-tetraloop and cutd2-tetraloop. Apparently cutd2-tetraloop has significant higher RMSDs for all residue, loop and stem. This suggests a less stable structure relative to both the standard tetraloop and the d2 tetraloop. Additionally 805-standard tetraloop had a slightly higher RMSD than 314-d2-tetraloop. This is consistent with data mining results from Williams group.

Table 3.1: All residues, stem and loop RMSD statistics for 805-standard tetraloop, 314-d2-tetraloop and cutd2-tetraloop.

| | | 805-standard tetraloop | 314-d2-tetraloop | cutd2-tetraloop |
|---|---|---|---|---|
| **All Residue** | MEAN | 2.333269036 | 1.525487289 | 4.241375335 |
| | STDEV | 0.376712833 | 0.265605061 | 0.699472294 |
| **Loop** | MEAN | 1.191101767 | 0.961592077 | 4.52860006 |
| | STDEV | 0.26507096 | 0.22334052 | 0.758655222 |
| **Stem** | MEAN | 1.583449308 | 1.38375446 | 4.073410149 |
| | STDEV | 0.321085543 | 0.158898361 | 0.708248029 |

The RMSDs of the terminus base pair, the central base pair and the closing base pair of 314-d2-tetraloop were calculated over the simulation time and were plotted on the same graph with the calculated RMSDs of each base pair of cutd2-tetraloop for stability comparison (Figure 3.7, Figure 3.8 and Figure 3.9). The comparison was very similar to the one conducted in Chapter II (between the standard tetraloop and the perfect d2 tetraloop) and showed us once again that the perfect d2 tetraloop is not stable relative to the d2 tetraloop with insertion and the standard tetraloop. All three stem base pairs RMSD statistics for 805-standard tetraloop, 314-d2-tetraloop and cutd2-tetraloop were summarized in Table 3.2.

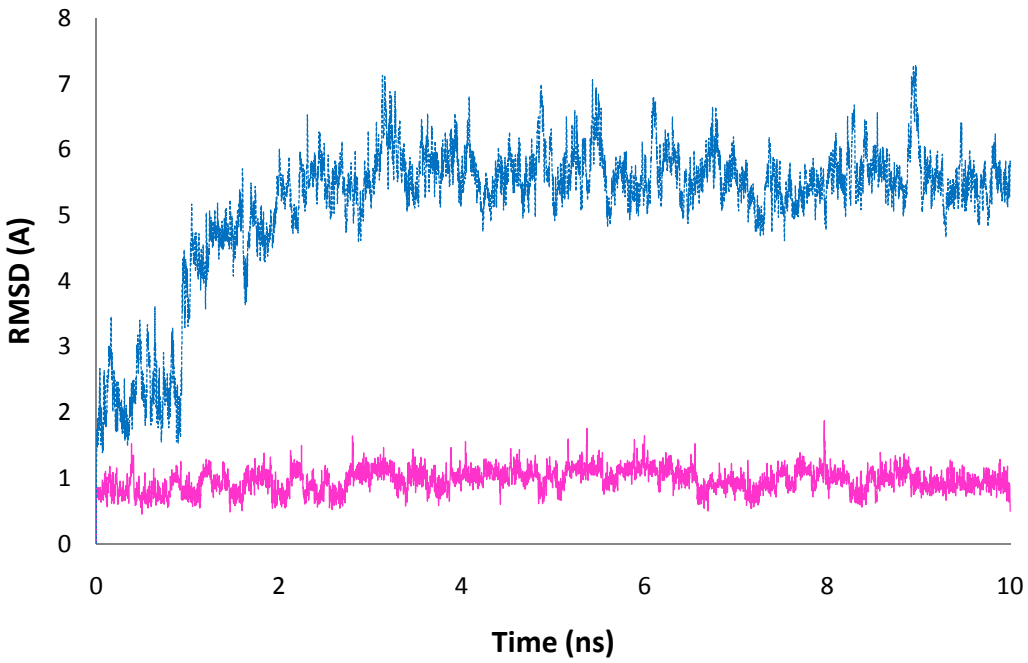



Figure 3.7: The terminus base pair RMSDs calculated over 10 ns of simulation for both the d2 tetraloop (314-d2tetraloop in solid pink line) and the perfect d2 tetraloop (cutd2-tetraloop in dotted blue line).

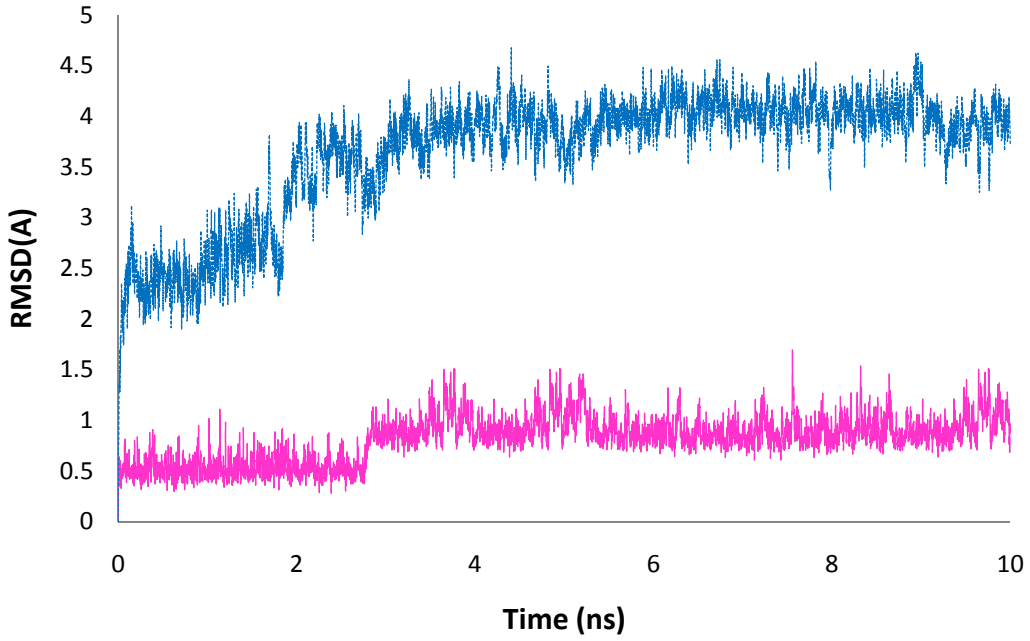



Figure 3.8: The central base pair RMSDs calculated over 10 ns of simulation for both d2 tetraloop with bulge (314-d2tetraloop in solid pink line) and perfect d2 tetraloop (cutd2-tetraloop in dotted blue line).

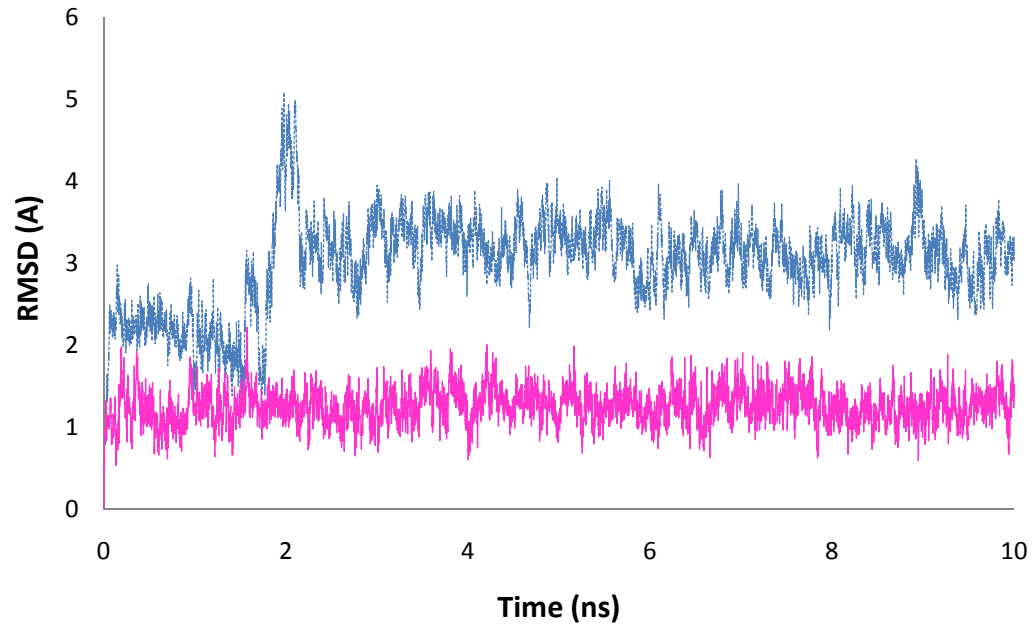



Figure 3.9: The closing base pair RMSDs calculated over 10 ns of simulation for both d2 tetraloop with bulge (314-d2tetraloop in solid pink line) and perfect d2 tetraloop (cutd2-tetraloop in dotted blue line).

Table 3.2: All three stem base pairs RMSD statistics for 805-standard tetraloop, 314-d2-tetraloop and cutd2-tetraloop.

| | | 805-standard tetraloop | 314-d2-tetraloop | cutd2-tetraloop |
|---|---|---|---|---|
| **Terminus** | MEAN | 1.546586 | 0.966021 | 5.184926 |
| **BP** | STDEV | 0.338601 | 0.175606 | 1.068514 |
| **Central** | MEAN | 1.234556 | 0.813275 | 3.646429 |
| **BP** | STDEV | 0.254065 | 0.225512 | 0.61418 |
| **Closing** | MEAN | 1.740621 | 1.252121 | 3.016755 |
| **BP** | STDEV | 0.224472 | 0.218351 | 0.555484 |

Table 3.2 recorded the RMSD statistics of all three base pairs in the stem for 805-standard tetraloop, 314-d2-tetraloop and cutd2-tetraloop. Again cutd2-tetraloop had significantly higher RMSD with large standard deviation for all three stem base pairs compared to 805-standard tetraloop and 314-d2-tetraloop. 805-standard tetraloop had slightly higher RMSDs compared to 314-d2-tetraloop, which was consistent with RMSDs comparisons in Table 3.1.

In sum, the RMSD analyses showed cutd2-tetraloop was not stable compared to 314-d2-tetraloop. The difference between the two tetraloop structures was the insertion residue was removed from the stem of cutd2-tetraloop. This means that the insertion residue plays an important role in stabilizing the structure of 314-d2-tetraloop and validates our hypothesis II.

*3.3.1.3 Hydrogen Bond and Base Pairing Tracking in 314-d2-tetraloop.*

Similar to the hydrogen bonding calculation conducted in Chapter II, we mainly focus our discussion on the conservation of the hydrogen bonding of those two characterizing intra-loop interactions (j-1–j+1 and j-1–j+2) and the hydrogen bonding interactions between each base pair in the stem during the simulation for 314-d2-tetraloop for stability comparison with cutd2-tetraloop.

*3.3.1.3 a) Stem Base Pairing*

The distance between the proton donor and acceptor of the terminus base pair, the central base pair and the closing base pair in the stem of 314-d2-tetraloop were calculated and plotted in Figure 3.10, Figure 3.11 and Figure 3.12 respectively to track the stem base pairing during the simulation.

Figure 3.10, Figure 3.11 and Figure 3.12 showed that all the hydrogen bonding interactions in each of the base pair in the stem of 314-d2-tetraloop were conserved perfectly throughout 10 ns of MD simulation. In contrast, the hydrogen bonding interactions in both the terminus and the central base pairs of cutd2-tetraloop were lost soon after the simulation started (Figure 2.20 and Figure 2.21), only the hydrogen bonding in the closing base pair of cutd2-tetraloop was conserved during the simulation (Figure 2.22). This Again suggested the perfect d2 tetraloop was less stable relative to the d2 tetraloop with insertion in the stem.

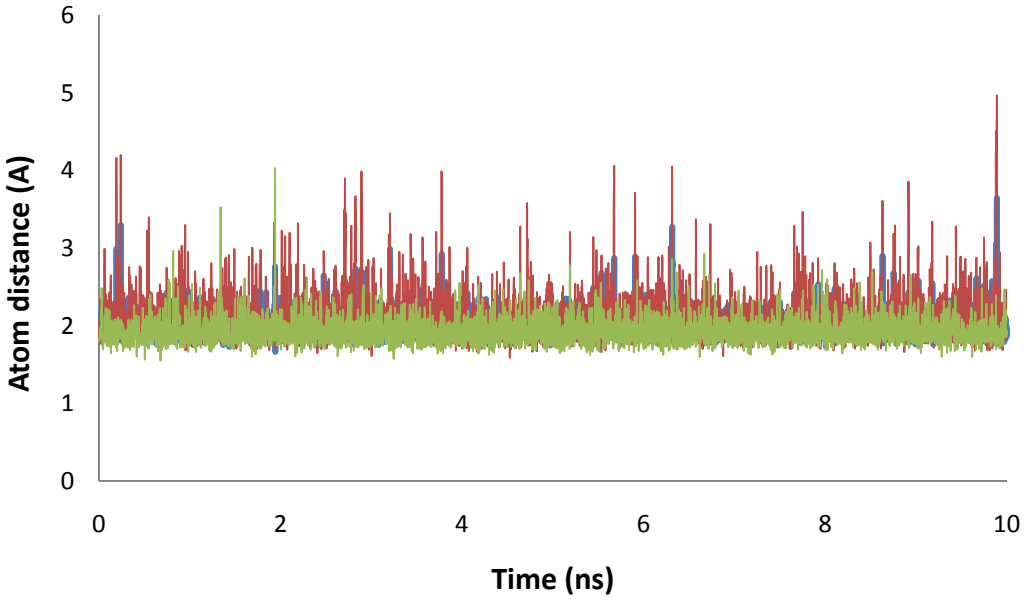



Figure 3.10: The terminus base pair of 314-d2-tetraloop. Green line is the atom distance between O2 on 311 Cytosine and H21 on 320 Guanine. Red line is the atom distance between H41 on 311 Cytosine and O6 on 320 Guanine. Blue line is the atom distance between N3 on 311 Cytosine and H1 on 320 Guanine.

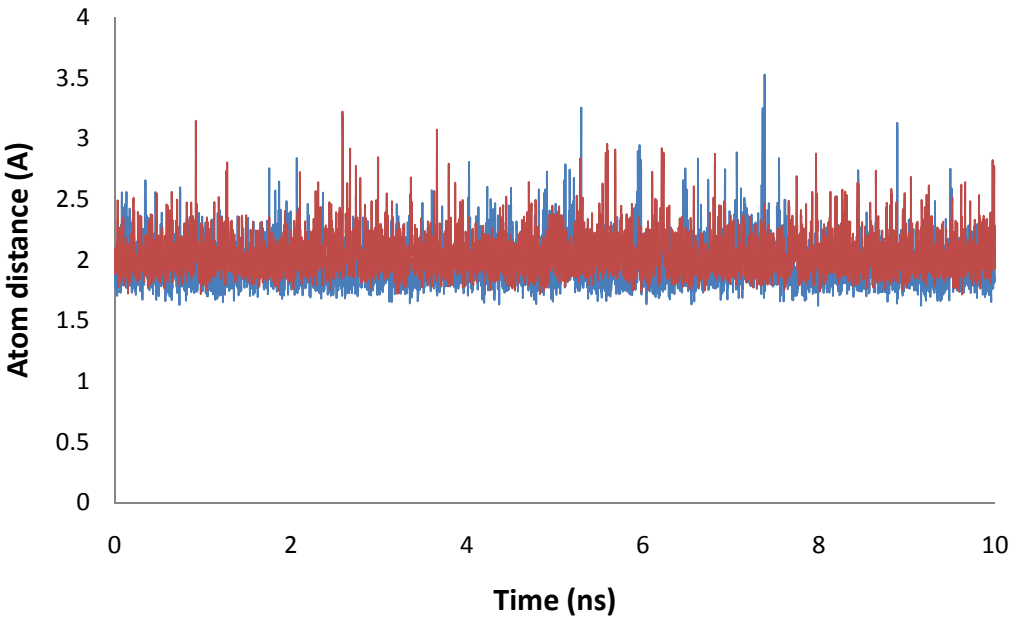



Figure 3.11: The central base pair of 314-d2-tetraloop. Red line is the atom distance between H3 on 312 Uracil and N1 on 319 Adenine. Blue line is the atom distance between O4 on 312 Uracil and H61 on 319 Adenine.

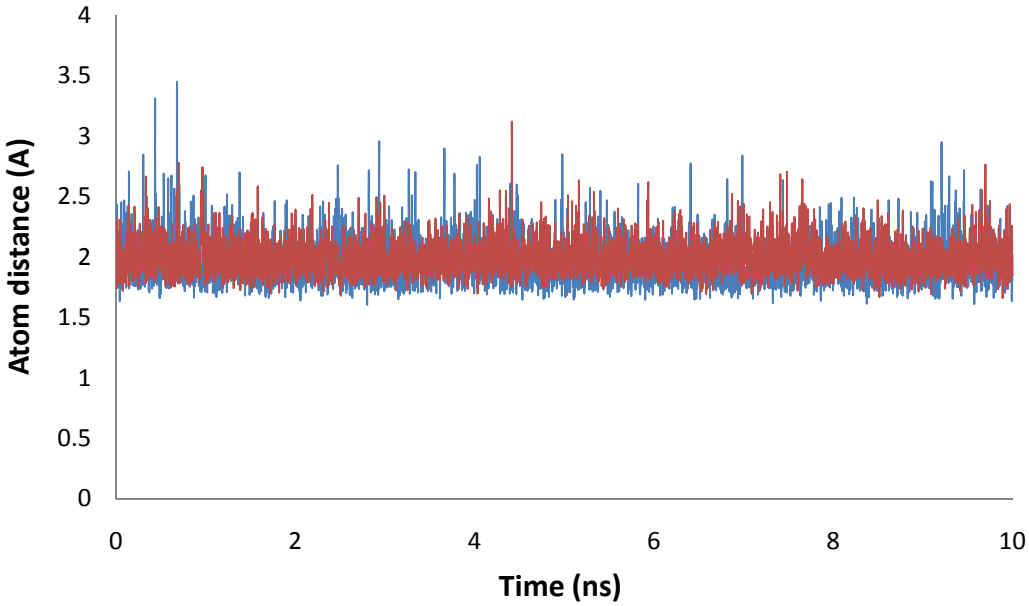



Figure 3.12: The closing base pair of 314-d2-tetraloop. Red line is the atom distance between H3 on 313 Uracil and N7 on 317 Adenine. Blue line is the atom distance between O4 on 313 Uracil and H61 on 317 Adenine.

Table 3.3 summarized the atom distance statistics indicating the hydrogen bonding interactions for each base pair in the stem of 314-d2-tetraloop. For the terminus base pair formed three hydrogen bonding interactions 311C(O2)–320G(H21), 311C(H41)–320G(O6) and 311C(N3)–320G(H1), therefore labelled Hbond1, Hbond2 and Hbond3 in the table. The central base pair and closing base pair both form two hydrogen bonding interactions: 312U(O4)–319A(H61) and 312U(H3)–319A(N1), and 313U(O4)–317A(H61) and 313U(H3)–317A(N7), both labeled as Hbond1 and Hbond2 in the table. All distances were averaged around 2 Å with very small standard deviation. The hydrogen bonding of each stem base pair showed a very stable stem of the d2 tetraloop with insertion.

Table 3.3: 314-d2-tetraloop hydrogen bonding statistics for each base pair in the stem.

| | | Terminus BP | Central BP | Closing BP |
|---|---|---|---|---|
| **Hbond1** | Mean | 1.928738 | 2.037116 | 1.990491 |
| | Stdev | 0.159014 | 0.157743 | 0.13779 |
| **Hbond2** | Mean | 3.087158 | 1.955503 | 1.990491 |
| | Stdev | 1.094979 | 0.175458 | 0.13779 |
| **Hbond3** | Mean | 2.097864 | | |
| | Stdev | 0.286994 | | |

*3.3.1.3 b) Intra-loop base pairing*

In the d2 tetraloop, the characterizing intra-loop hydrogen bonding interactions are the one between hydroxyl group on the sugar ring of residue j-1 and cross-loop base atoms of residue j+1, and the one between base atoms of residue j-1 and backbone atom O2P of residue j+2. Since residue j+2 of the d2 tetraloop is shifted down to position j+3 of the standard tetraloop, there is no hydrogen bonding interaction between the base of residue j-1 and the base of residue j+2.

To compare the loop stability of the d2 tetraloop and the perfect d2 tetraloop, these two hydrogen bonding interactions of 314-d2-tetraloop during the simulation were calculated and plotted on the same graph with those of cutd2-tetraloop (Figure 3.13 and Figure 3.14).

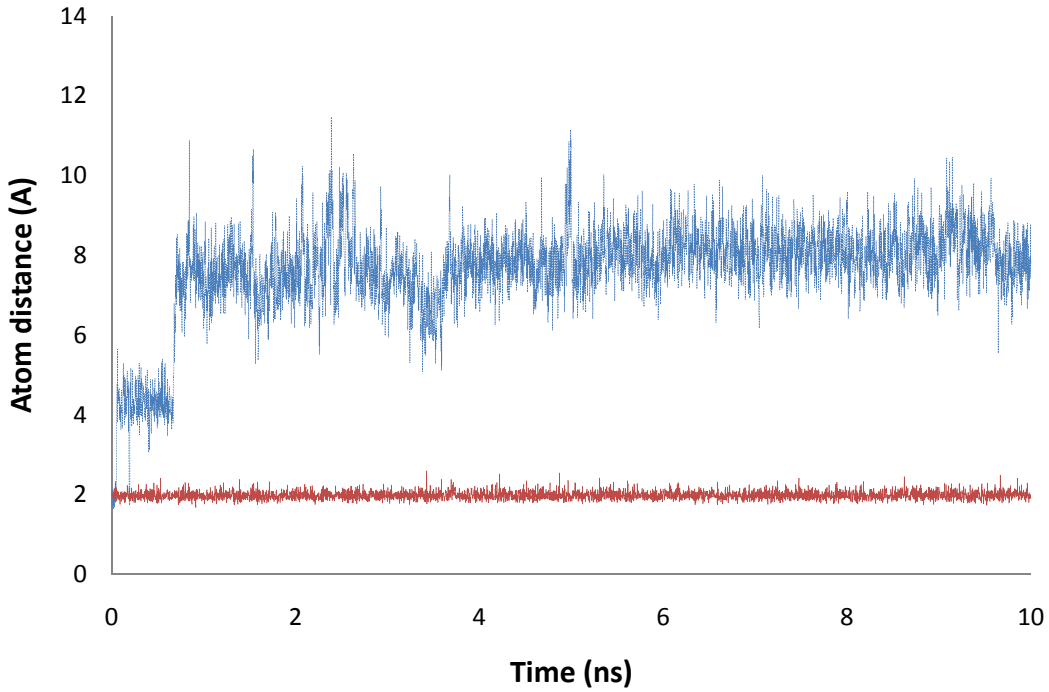



Figure 3.13: Intra-loop hydrogen bonding interaction (j-1 and j+1) comparison between 314-d2-tetraloop (Solid red line) and cutd2-tetraloop (dotted blue line). The distance between H2' on Guanine of residue 314 at position j-1 and N7 on Adenine of residue 316 at position j+1 for both 314-d2-tetraloop and cutd2-tetraloop.

Similar to the observed comparison of the intra-loop hydrogen bonding interactions (j-1–j+1) between the standard tetraloop and the d2 tetraloop (Figure 2.23), Figure 3.13 showed that Guanine of residue 314 (position j-1) was forming hydrogen bonding with Adenine of residue 316 (position j+1) throughout 10 ns of the simulation. The average distance between H2' of Guanine on residue 314 and N7 of Adenine on residue 316 is 2.05 Å with a standard deviation of 0.51 Å. On the contrary, the hydrogen bonding interaction between residue j-1 and residue j+1 of cutd2-tetraloop was not conserved during the simulation.

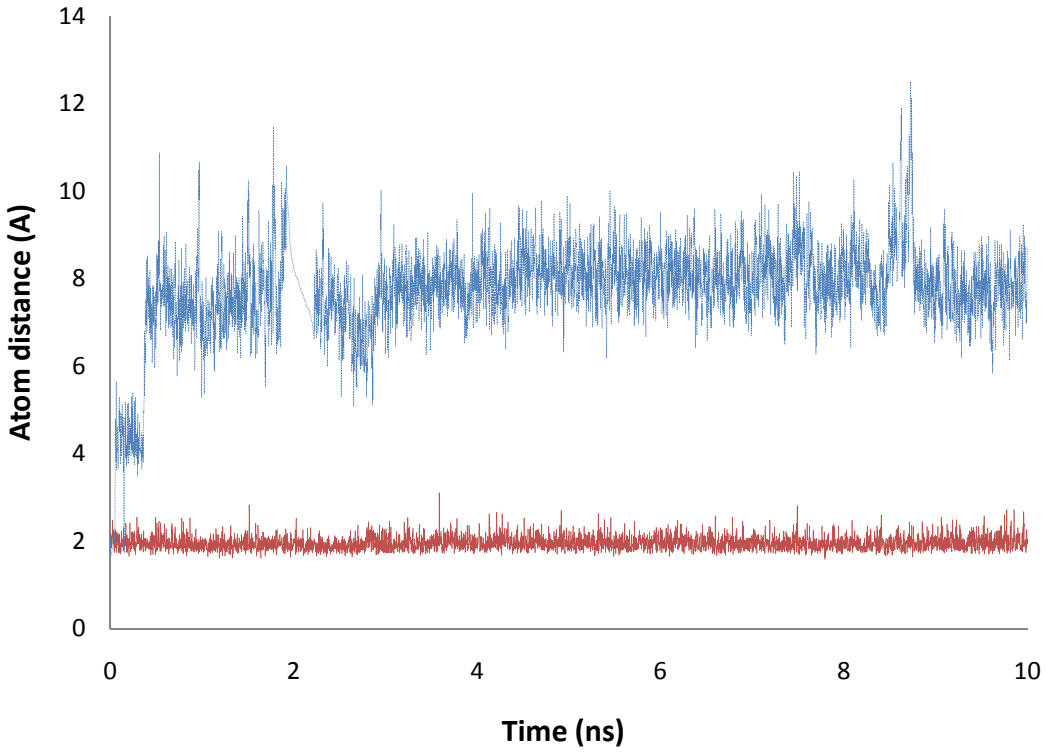



Figure 3.14: Intra-loop hydrogen bonding interaction (j-1 and j+2) comparison between 314-d2-tetraloop (red solid line) and cutd2-tetraloop (blue dotted line). The distance between H1 on Guanine of residue 314 at position j-1 and O2P on Adenine of residue 317 at position j+2 for both 314-d2-tetraloop and cutd2-tetraloop.

Figure 3.14 showed that Guanine of residue 314 (position j-1) was forming hydrogen bonding with Adenine of residue 317 (position j+2) throughout 10 ns of the simulation. The average distance between H1 of Guanine on residue 314 and O2P of Adenine on residue 317 is 2.07 Å with a standard deviation of 0.35 Å. In contrast, the hydrogen bonding interaction between residue j-1 and residue j+1 of cutd2-tetraloop was not conserved during the simulation.

In sum, calculations of intra-loop hydrogen bonding interactions (the one between j-1 and j+1 and the one between j-1 and j+2) from 314-d2-tetraloop simulation showed consistency with data mining results. Comparisons of these intra-loop hydrogen bonding interactions between 314-d2-tetraloop and cutd2-tetraloop again showed that the perfect d2 tetraloop was not a stable structure compared to the d2 tetraloop.

*3.3.1.4 Torsion Angle Analysis of 314-d2-tetraloop*

The same torsion angle analysis was conducted on the simulation of the d2 tetraloop (314-d2-tetraloop) as the one conducted on the simulations of the standard tetraloop (805-standard tetraloop) and the perfect d2 tetraloop (cutd2-tetraloop) in Chapter II.

Ten ns of Molecular Dynamic simulation of 314-d2-tetraloop showed a very stable structure of the d2 tetraloop in terms of backbone torsion angle distribution. Four backbone torsion angles (α, γ, δ, and ζ) of the loop region residues throughout the simulation were measured and recorded in Table 3.4. For each torsion angle the first row is torsion angle value of the starting structure, the second row is the average value of the torsion angle during the simulation, and the third row is the standard deviation. Three columns represent torsion angles of three residues 314G j-1, 315G residue j, and 316A residue j+1 in the loop region.

Table 3.4: Four backbone torsion angle for three residues in the loop of 314-d2-tetraloop

| | | 314G | 315G | 316A |
|---|---|---|---|---|
| α | Starting Point | 290.60 | 175.71 | 292.04 |
| | Mean | 289.92 | 78.26 | 290.81 |
| | Stdev | 13.24 | 33.58 | 21.55 |
| γ | Starting Point | 55.75 | 54.52 | 67.85 |
| | Mean | 54.07 | 179.44 | 55.84 |
| | Stdev | 9.92 | 32.32 | 11.21 |
| δ | Starting Point | 78.40 | 85.38 | 87.44 |
| | Mean | 74.74 | 82.82 | 83.92 |
| | Stdev | 7.12 | 11.98 | 7.21 |
| ζ | Starting Point | 295.23 | 279.30 | 289.27 |
| | Mean | 264.60 | 281.70 | 304.00 |
| | Stdev | 13.11 | 58.70 | 12.66 |

The binning of ASCII code for 314-d2-tetraloop is 'aoa' [7], which represents bin number 3111, 2111, and 3111 respectively. According to bin assignments and torsion angles (Table 2.1), 3111 means α with range of 260° – 330°, γ with range of 35° – 75°, δ with range of 68° – 93°, and ζ with range of 255° – 325°. And 2111 means α with range of 135° – 190°, γ with range of 35° – 75°, δ with range of 68° – 93°, and ζ with range of 255° – 325°.

Backbone torsion angle analysis of 314-d2-tetraloop from the simulation showed again that the d2 tetraloop is a very stable structure. Most of the backbone torsion angles calculations from the simulation were consistent with the empirical

observations. Only backbone torsion angle α and γ of residue 315G were not consistent with empirical observation.

In contrast, 10 ns of Molecular Dynamics simulation of cutd2-tetraloop showed much less stable structure of perfect d2 tetraloop with no defects in the stem. Four backbone torsion angles (α, γ, δ, and ζ) of loop region residues throughout the simulation were also measured and recorded in Table 2.5, which showed large torsion angle fluctuations.

The comparison of backbone torsion angle analysis between cutd2-tetraloop and 314-d2-tetraloop from the MD simulation showed again that the perfect d2 tetraloop with no stem defects is not a stable structure relative to the d2 tetraloop. The average value of the backbone torsion angles of three residues in the loop of cutd2-tetraloop fell outside of the range suggested by the binning method and had very large standard deviation. The backbone torsion angle distribution of 314-d2-tetraloop is similar to that of 805-standard tetraloop, which suggested a stable structure of both the standard tetraloop and the d2 tetraloop.

In sum, RMSDs calculations, hydrogen bonding calculations, and torsion angle analyses of 314-d2-tetraloop showed a stable d2 tetraloop with insertion.

### 3.3.2 Hydrogen Bonding of Insertion Residue 318C

*3.3.2.1 Observed Hydrogen Bonding of Insertion with both the Loop and the Stem*

The MD simulation of 314-d2-tetraloop revealed an interesting phenomenon. It was observed that during the simulation the insertion residue was involved in two hydrogen bonding interactions. First, the base Cytosine (318C) on the insertion residue was interacting with the hydroxyl group on the sugar ring of residue 316A at position j+1 in the loop region, forming a hydrogen bond. The other hydrogen bond was formed between the hydroxyl group on the sugar ring of insertion residue 318C and the base Adenine of residue 317A in the closing base pair. Both hydrogen bonding interactions

were highlighted in red dotted line in Figure 3.15. Therefore the insertion residue 318C appears to form hydrogen bonding interactions with both the loop and the stem in 314-d2-tetraloop.

These hydrogen bonding interactions formed by the insertion residue with both the loop and the stem of 314-d2-tetraloop could explain how stem defects, an insertion residue in the stem in this case, stabilize a structure like the d2 tetraloop. We originally believed that the stem imperfection, an insertion residue in the case of 314-d2-tetraloop, somehow stabilizes the d2-tetraloop by forming a conformationally more relaxed stem, Our simulation showed hydrogen bonding formed both between the insertion residue and the loop, and between the insertion residue and the stem. As a result, it appeared that the insertion (or stem imperfection) could also stabilize the d2-tetraloop by decreasing the enthalpy of the system. The simulation comparison of 314-d2-tetraloop and cutd2-tetraloop showed that the insertion residue is the stabilizing factor of 314-d2-tetraloop and the hydrogen bonding of insertion residue showed how it stabilizes 314-d2-tetraloop.

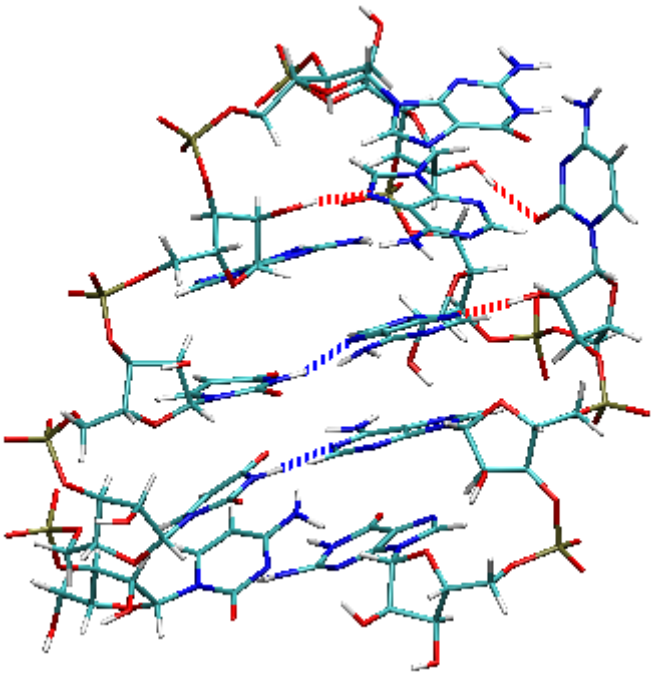


Figure 3.15: Snapshot of 314 d2-tetraloop from the simulation with hydrogen bond highlighted in dotted line.

*3.3.2.2 Hydrogen bonding calculation of insertion residue 318C*

MD simulation of 314-d2-tetraloop exhibited that the hydroxyl group on the sugar ring of residue 316A in the loop region was forming multiple hydrogen bonding interactions with multiple proton acceptors on the base cytosine of insertion residue 318C, such as N3, C2 and O2. The atom distances between H2' of residue 316A and all possible proton acceptors of residue 318C were calculated in Figure 3.16.

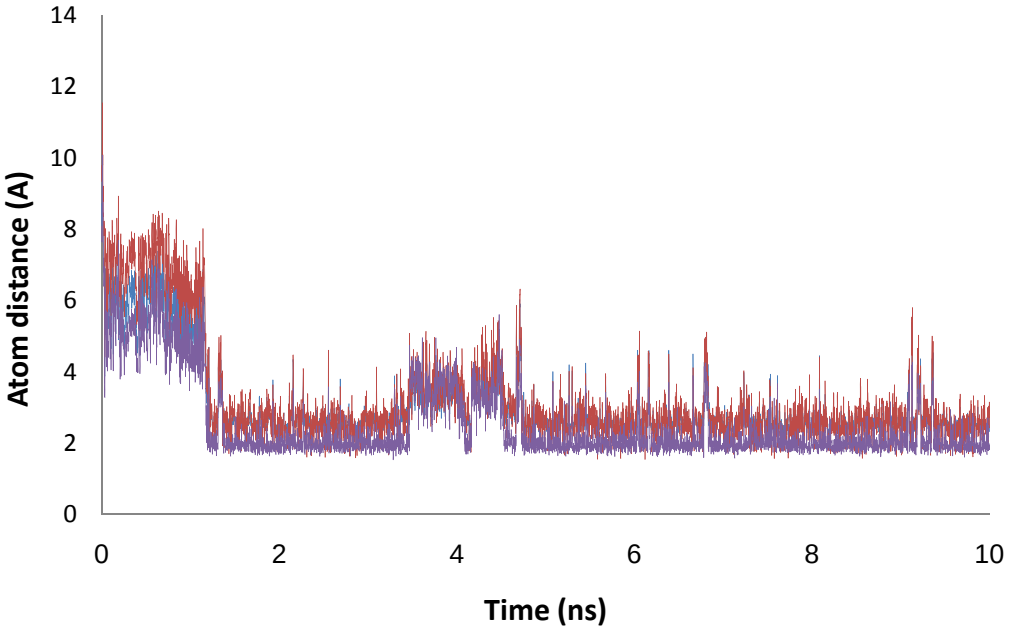



Figure 3.16: Hydrogen bonding interactions between residue H2' of 316A (j+1) and cytosine of 318C, atom distance between 316A(H2') and 318C(C2) in blue, 316A(H2') and 318C(O2) in red, and 316A(H2') and 318C(N3) in purple.

Other than the hydrogen bonding interactions with the loop region, the insertion residue was observed to form hydrogen bonding with the residue in the stem of 314-d2-tetraloop as well during the simulation. The hydroxyl group of insertion

residue 318C was forming hydrogen bonding interaction with the base of residue 317A in the stem. The atom distance between N3 of residue 317A and H2' of residue 318C was calculated and plotted in Figure 3.17. From Figure 3.17 we can see that the hydrogen bonding interaction between the insertion residue and the stem was conserved for most of the time.

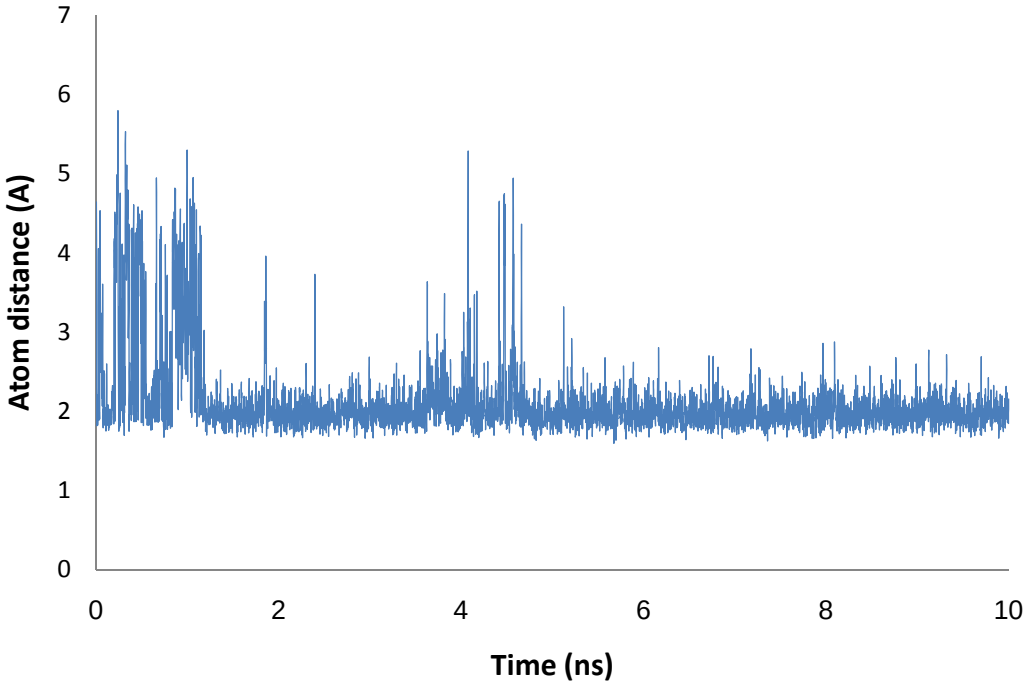



Figure 3.17: Hydrogen bond interaction between hydroxyl group of insertion residue 318C and the base of residue 317A in the stem, atom distance between 318C(H2') and 317A( N3).

The hydrogen bonding interactions of insertion residue 318C with both the stem and the loop appear to be contributing to the stability of the structure by lowering the system enthalpy. It is worth further investigating on whether conformational relaxation of the stem by the insertion residue also plays a role in contributing to the stability of the d2 tetraloop. One way to test this is to remove the enthalpy contribution

by eliminating the hydrogen bonding between insertion residue with both the loop and the stem. This was conducted in the following sequential order:

1. Remove the hydrogen bonding interactions between the insertion residue and the loop by deoxidizing the sugar ring of residue 316A. Run 10ns of MD simulation at 310K on the modified structure.
2. Remove the hydrogen bonding interactions of the insertion residue with both the loop and the stem by deoxidizing the sugar ring of both residue 316A and residue 318C. Run 10ns of MD simulation at 310K on the modified structure.

### 3.3.3 Simulations of de316-314-d2-tetraloop

As described in step I of deoxidization, the hydroxyl group was removed from the sugar ring of residue 316A to eliminate the hydrogen bonding interactions between the loop and the insertion (Figure 3.18). The structure on the left is 314-d2-tetraloop. Two dash lines were used to represent the two hydrogen bonding interactions: the one between the insertion and the loop, and the one between the insertion and the stem. The structure on the right is the modified one where 316A was deoxidized (hydroxyl removed compared to 314-d2-tetraloop on the left). The new structure was labeled de316-314-d2-tetraloop. This way the dash line representing the hydrogen bonding interaction between the insertion and the loop in 314-d2-tetraloop was gone in de316-314-d2-tetraloop.

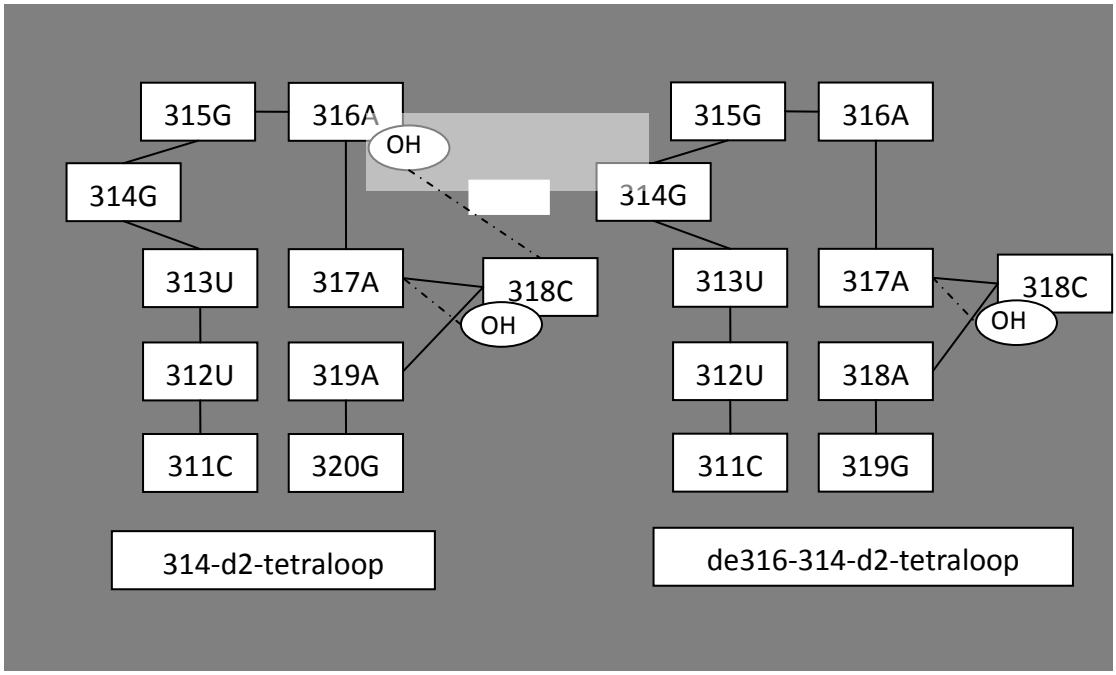



Figure 3.18: Modification of 314-d2-tetraloop by removing the hydroxyl group (2') from the sugar ring of residue 316A. The structure on the left represents the original 314-d2-tetraloop with two hydrogen bonding of the insertion residue 318C (dash line). The structure on the right represents the deoxidized 314-d2-tetraloop, labeled de316-314-d2-tetraloop with only the hydrogen bonding between insertion residue 318C and the stem remaining.

*3.3.3.1 Simulation observation*

MD simulation of de316-314-d2-tetraloop showed that the hydrogen bonding between insertion residue 318C and the loop is important for stabilizing the loop structure. Although the sugar ring of residue 316A was deoxidized, the loop was still attempting to form hydrogen bonding with the insertion residue 318C (dotted lines highlighted the hydrogen bonding interactions on the left structure in Figure 3.19), such as 316A(N3) – 318C(H42), 316A(N3) – 318C(H41), and 316A(H21') – 318C(N4). However, these hydrogen bond interactions did not last long and were broken half way through the simulation.

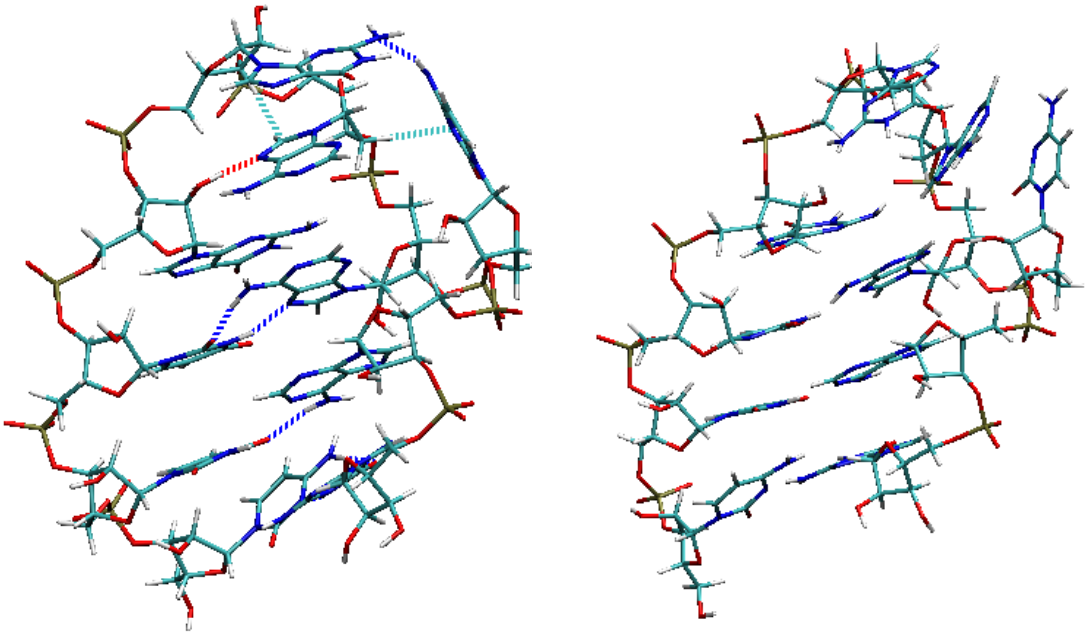


Figure 3.19: Snapshots of de316-314-d2-tetraloop at 1.5ns of the simulation (left), which illustrated that the insertion was attempting to form hydrogen bonding with the loop, and at the end of the simulation (right).

The end structure of de316-314-d2-tetraloop from 10ns of simulation was exhibited in Figure 3.19. Compared to the structure of the perfect d2 tetraloop (cutd2-tetraloop) where the insertion residue 318C was removed completely, the structure of deoxidized 314-d2-tetraloop (de316-314-d2-tetraloop) remained relatively stable during the simulation, especially the stem where nice base pairing and stacking were conserved throughout the simulation. However the intra-loop hydrogen bonding was not conserved very well and the loop structure was observed to experience atomistic changes (detailed RMSD and hydrogen bonding calculations in the next section).

*3.3.3.2 RMSD Analysis of de316-314-d2-tetraloop*

The simulation trajectory of de316-314-d2-tetraloop was superposed on the starting structure before the RMSD values were calculated. All atoms, loop and stem RMSD (Figure 3.20 and Figure 3.21) from the simulation showed a relatively stable structure of de316-314-d2-tetraloop compared to those of cutd2-tetraloop where the insertion residue was removed completely from the stem. The RMSDs also showed that de316-314-d2-tetraloop appeared to have experienced some structural changes after about 5 ns of the simulation. The all atoms RMSDs have a mean of 1.53 Å and a standard deviation of 0.34 Å before 5ns of the simulation. After 5ns the mean and standard deviation of all atoms RMSD become 2.51 Å and 0.44 Å.

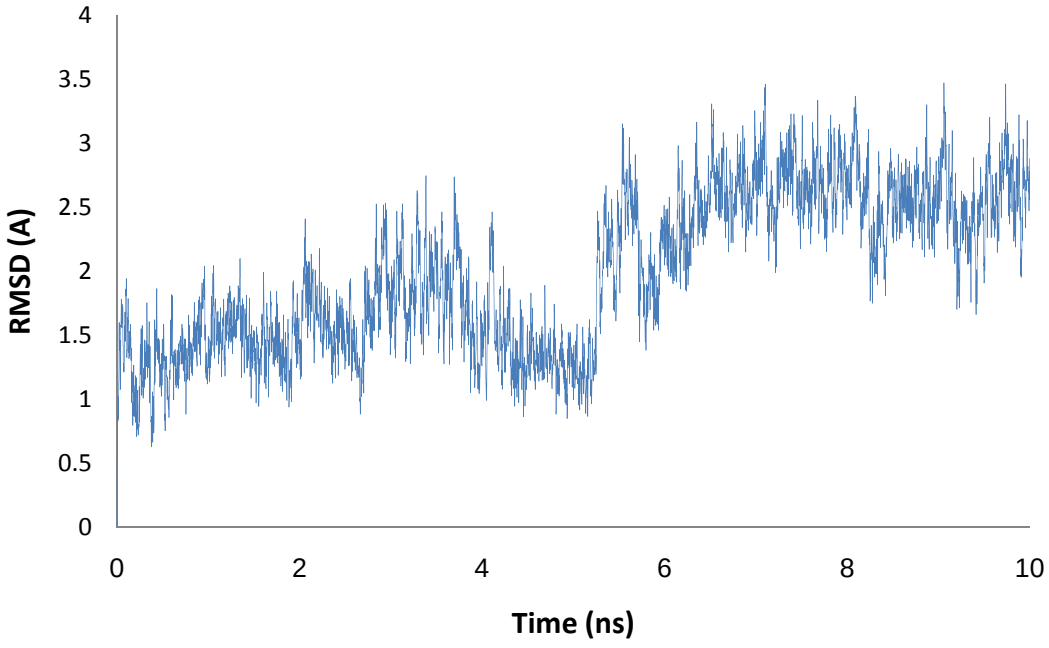



Figure 3.20: All atoms RMSD calculations of de316-314-d2-tetraloop over 10ns of simulation time.

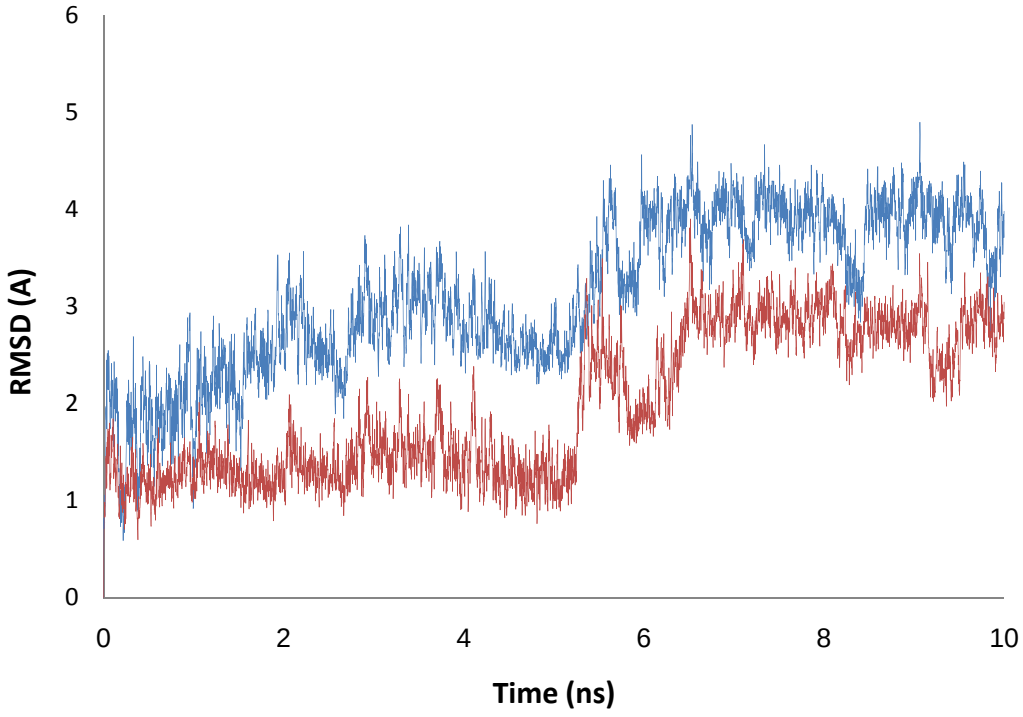



Figure 3.21: Loop (blue) and stem (red) region RMSD calculations of de316-314-d2-tetraloop over 10ns of simulation time.

The loop and the stem region RMSDs of de316-314-d2-tetraloop were both calculated and plotted in the same graph for comparison (Figure 3.21). The comparison showed that the loop region of de316-314-d2-tetraloop was experiencing larger structural deviation than the stem region. The average RMSD of the loop is 3.12 Å while the average RMSD of stem is 1.96 Å. The RMSD values of the loop and the stem both showed de316-314-d2-tetraloop were experiencing a slightly structural change at around 5ns of the simulation as did all atoms RMSD values. This slightly change of structure could be further explained by the hydrogen bonding analysis conducted in the next section.

In sum, the RMSD analysis of de316-314-d2-tetraloop showed a relatively stable structure compared to the RMSD analysis of cutd2-tetraloop conducted in Chapter II. The structure of de316-314-d2-tetraloop appeared to deviate from the

structure of 314-d2-tetraloop when the hydroxyl group on the sugar ring of residue 316A was removed. The loop of de316-314-d2-tetraloop was experiencing more structural changes than the stem.

#### *3.3.3.3 Hydrogen Bonding Analysis of de316-314-d2-tetraloop*

From the observation of the simulation trajectory, we can see that even in the absence of the hydroxyl group on the sugar ring of residue 316A (j+1), the loop of de316-314-d2-tetraloop was still attempting to form hydrogen bonding interactions with the insertion residue 318C, such as 316A(N3) – 318C(H42), 316A(N3) – 318C(H41), and 316A(H21') – 318C(N4), which were highlighted in dotted line in the left snapshot of Figure 3.19.

However, these hydrogen bond interactions did not last long and were broken half way through the simulation, which is probably the reason for the changing structure captured in the RMSD analysis after 5ns of the simulation. Three hydrogen bonding interactions between the loop and the insertion residue observed during the simulation, 316A(N3) – 318C(H42), 316A(N3) – 318C(H41), and 316A(H21') – 318C(N4), were calculated and plotted in Figure 3.22.

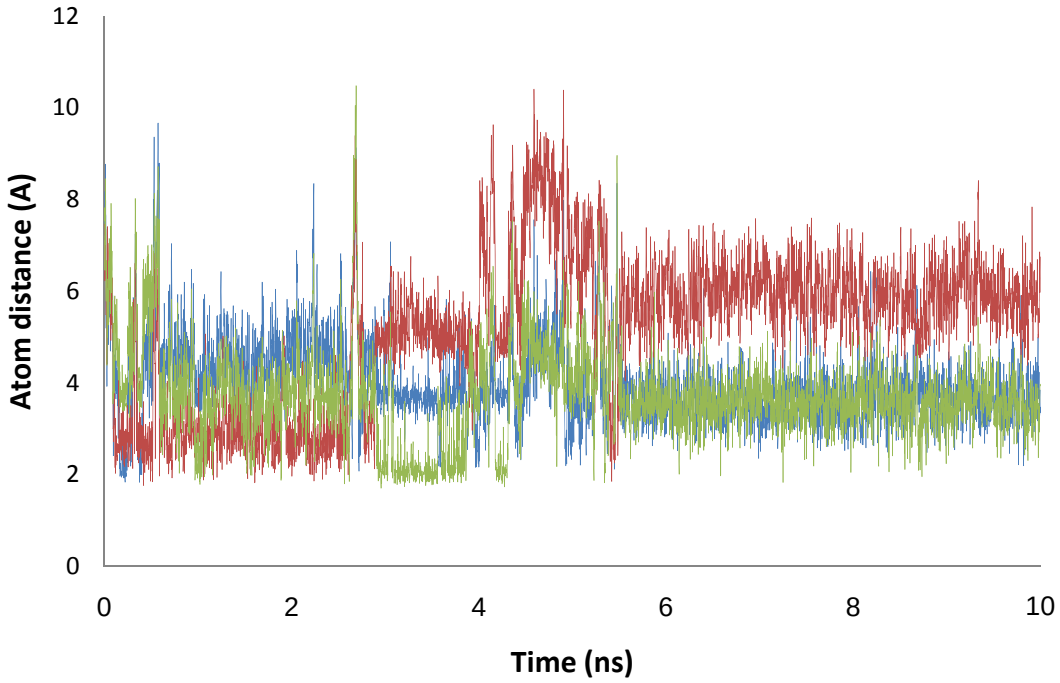



Figure 3.22: Distance calculated between atoms in the loop and the insertion residue that were observed to form hydrogen bonding in the simulation of de316-314-d2-tetraloop. Blue represents the atom distance between N3 on residue 316A and H41 on residue 318C, red represents the atom distance between H21' on residue 316A and N4 on residue 318C, and green represents the atom distance between N3 on residue 316A and H42 on residue 318C.

The atom distances between N3 on residue 316A and H41 on residue 318C, H21' on residue 316A and N4 on residue 318C, and N3 on residue 316A and H42 on residue 318C have a mean and a standard deviation of 3.82 Å and 1.02 Å, 5.78 Å and 1.63 Å, and 4.23 Å and 0.91 Å respectively, indicating that these hydrogen bonding interactions were not conserved during the simulation. All three atom distances experienced a significant increase after 5ns of simulation.

Figure 3.22 also exhibited that the hydrogen bonding interaction between 316A(H21') and 318C(N4) was conserved for the first 3ns and the hydrogen bonding interaction between 316A(N3) and 318C(H42) was conserved from 3ns to 4ns. Taken together, this suggested that the insertion residue 318C was forming hydrogen bonding

interactions with the loop during the first 4ns of the simulation. This was an important finding because the hydrogen bonding interaction between the insertion residue 318C and the loop appeared to be correlated with the intra-loop hydrogen bonding interaction between residue j-1 and residue j+1 over 10ns of simulation (Figure 3.23).

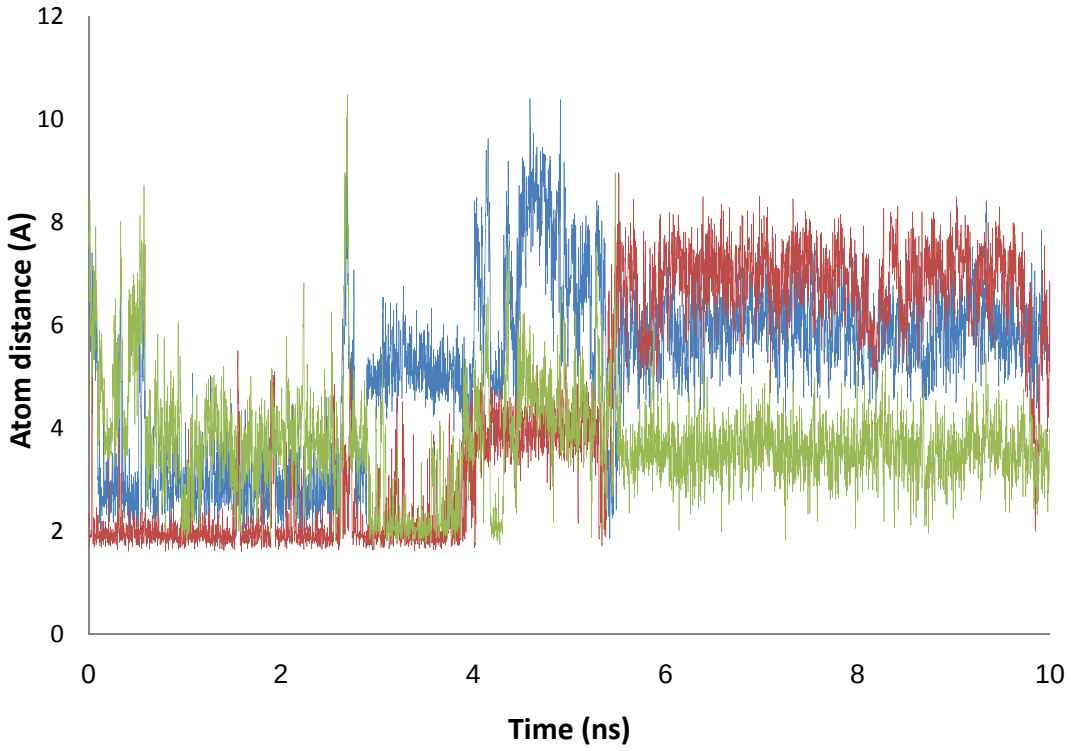



Figure 3.23: The intra-loop hydrogen bonding interaction of residue j-1 and residue j+1 (in red) was compared to the hydrogen bonding interaction between residue 316A in the loop and the insertion residue 318C (in blue) and the hydrogen bonding interaction between N3 on residue 316A and H42 on residue 318C.

The intra-loop hydrogen bond interaction between residue j-1 and residue j+1, was plotted on the same graph with the hydrogen bond interaction between H21' on residue 316A and N4 on residue 318C and the hydrogen bonding interaction between N3 on residue 316A and H42 on residue 318C (Figure 3.23).

From Figure 3.23 we can see that there is a correlation between the loss of loop-insertion hydrogen bonding and the loss of intra-loop hydrogen bonding (j-1 and j+1). The structure of the loop region of de316-314-d2-tetraloop deformed when the intra-loop hydrogen bonding interaction (j-1 and j+1) was broken. Therefore the hydrogen bond interaction between the loop and the insertion residue in the stem, or stem defects, appears to be a major factor for stabilizing the loop structure of the d2 tetraloop, 314-d2-tetraloop in this case.

However, the whole structure of de316-314-d2-tetraloop remained relatively stable throughout 10ns of the simulation especially when compared to the simulation results of the perfect d2 tetraloop without any defects in the stem from Chapter II (cutd2-tetraloop with insertion residue 318C removed completely from the stem of 314-d2-tetraloop). In addition, during the simulation the hydrogen bonding interaction between the insertion residue 318C and residue 317A in the stem of de316-314-d2-tetraloop was conserved for most of the simulation time (Figure 3.24). Figure 3.24 showed that the average atom distance between N3 on residue 317A and H2' on residue 318C was 2.49 Å. However the distance had a relatively large standard deviation 0.95 Å indicating a large fluctuation of this hydrogen bonding.

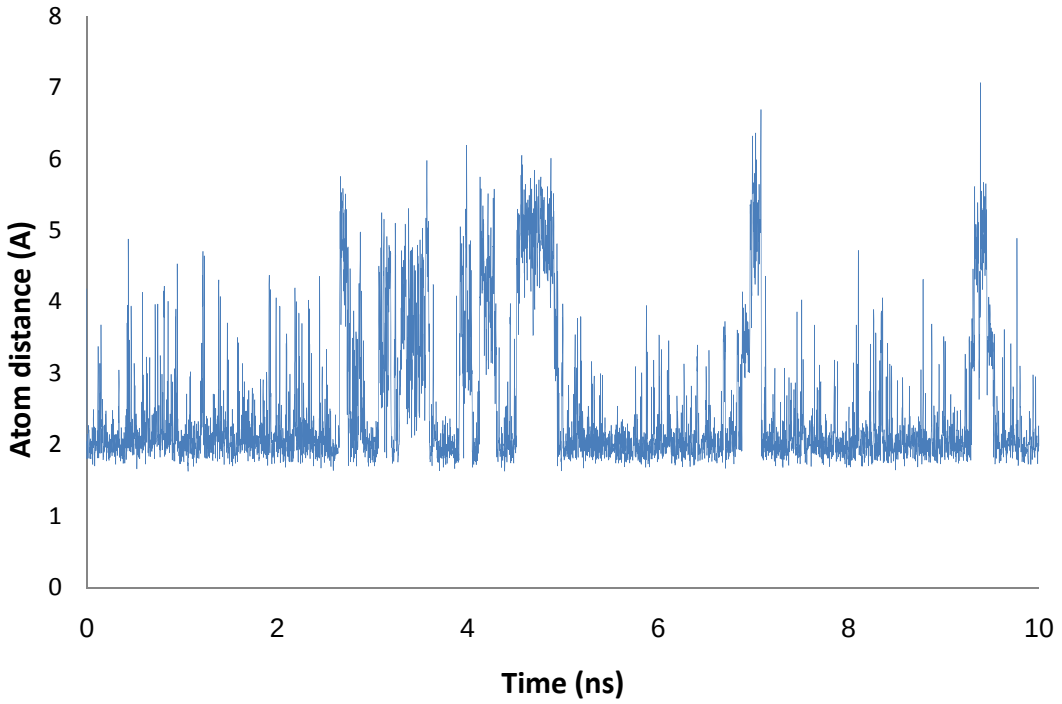



Figure 3.24: The hydrogen bonding interaction between N3 on residue 317A and H2' on residue 318C calculated over 10ns of the simulation.

In sum, when the hydroxyl group of residue 316A (j+1) in the loop was removed, the hydrogen bonding between the insertion residue and the loop region was eliminated. This made the structure of de316-314-d2-tetraloop less stable specifically in the loop region, which experienced atomistic changes and lost the intra-loop hydrogen bonding characterizing tetraloops.

The hydrogen bonding between the stem and the insertion residue was conserved relatively well. Could this stem-insertion hydrogen bonding interaction be the reason for holding the structure of de316-314-d2-tetraloop relatively stable or could the conformational relaxation of stem by the insertion contribute partly to the stability of the structure as well? To answer this question, the hydroxyl group on the insertion residue 318C was removed to eliminate the hydrogen bonding interaction between insertion and stem for further investigation.

### 3.3.4 Simulation Results of de316-de318-314-d2-tetraloop

As described in step II of deoxidization, the hydroxyl group was further removed from the sugar ring of residue 318C of de316-314-d2-tetraloop to eliminate the hydrogen bonding interaction between the stem and the insertion (Figure 3.25). The structure on the left is de316-314-d2-tetraloop, where the dash line represents the hydrogen bonding interaction between the stem and the insertion. The structure on the right is the modified one where 318C was deoxidized (hydroxyl group removed compared to de316-314-d2-tetraloop). The new structure on the right was labeled de316-de318-314-d2-tetraloop. In the new structure the dash line is gone which means the hydrogen bonding interactions between the stem and the insertion residue is eliminated.

This new structure, de316-de318-314-d2-tetraloop, was simulated at 300K for 10ns. RMSDs analysis and hydrogen bonding calculations were conducted after the simulation.

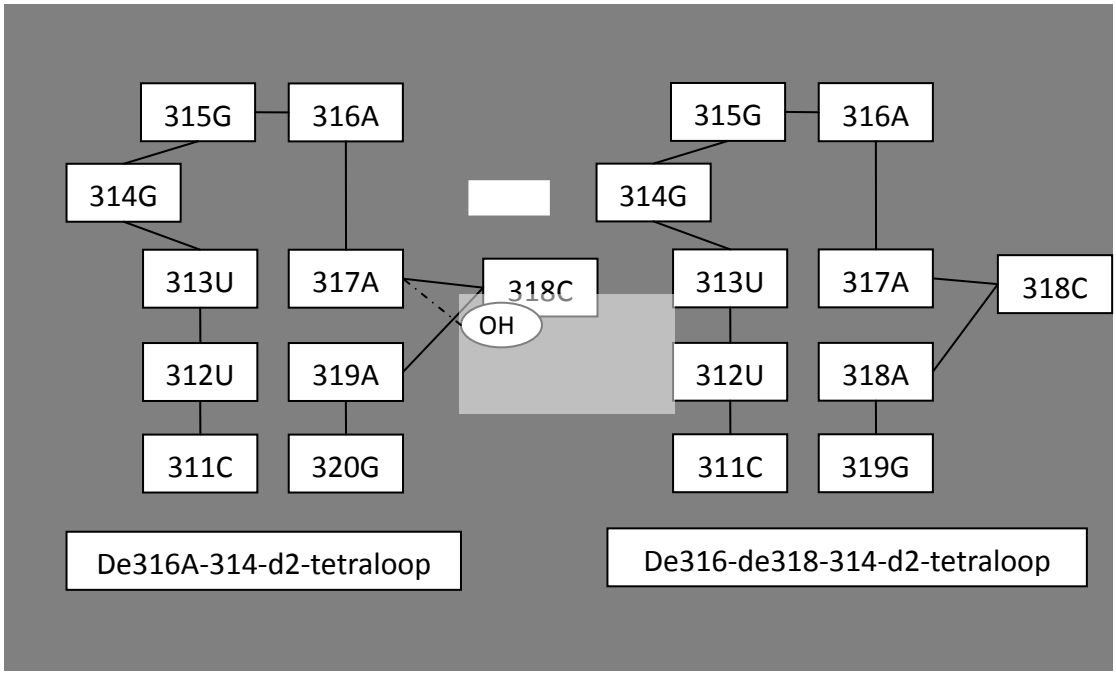



Figure 3.25: Modification of de316-314-d2-tetraloop by removing hydroxyl group (2') from sugar ring of residue 318C. The structure on the left represents the original de316-314-d2-tetraloop with the hydrogen bonding interaction between the insertion residue 318C and the stem. The structure on the right represents the deoxidized de316-314-d2-tetraloop, labeled de316-de318-314-d2-tetraloop with both hydrogen bonding of insertion residue 318C removed.

*3.3.4.1 Simulation Observation of de316-de318-314-d2-tetraloop*

Simulation observations of de316-de318-314-d2-tetraloop were very similar to those of de316-314-d2-tetraloop. The structure of the stem did not change appreciably even though the hydrogen bond interaction between the insertion residue 318C and residue 317A in the stem was eliminated. Note that although two nucleotides were deoxidized the starting atomistic positions of de316-de318-314-d2-tetraloop were the same as 314-d2-tetraloop. Unlike the perfect d2 tetraloop with no stem defects (cutd2-tetraloop) where the stem started to fall apart during the simulation, the stem of this de316-de318-314-d2-tetraloop with neither hydrogen bonding interactions of insertion residue (the one with the loop and the one with the stem) still had very nice

base pairing and stacking throughout the simulation. Snapshots of starting and ending structure of de316-de318-314-d2-tetraloop were exhibited in Figure 3.26.

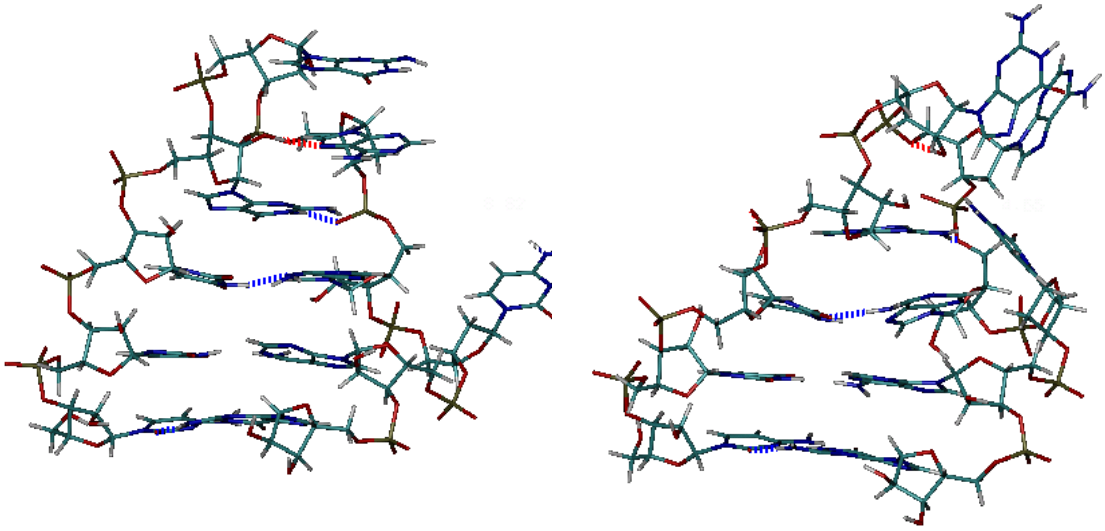


Figure 3.26: Snapshots of de316-de318-314-d2-tetraloop from 10ns of MD simulation. The left figure is the structure at the beginning of the simulation, and the right figure is the structure at the end of the simulation.

The loop structure of de316-de318-314-d2-tetraloop, however, did experience considerable changes from its starting point (Figure 3.26). What was different from the simulation of de316-314-d2-tetraloop was that those hydrogen bonding interactions between the loop and the insertion observed during the first 4ns of simulation of de316-314-d2-tetraloop could barely hold in the simulation of de316-de318-314-d2-tetraloop. In addition, the intra-loop hydrogen bonding interaction between residue j-1 and residue j+1 was broken soon after the simulation started, which was consistent with the correlation of intra-loop hydrogen bonding and loop-insertion hydrogen bonding observed during the first 4ns of the simulation of de316-314-d2-tetraloop.

Therefore Molecular Dynamics simulations of both de316-314-d2-tetraloop and de316-de318-314-d2-tetraloop showed that eliminating the loop-insertion hydrogen bonding and the stem-loop hydrogen bonding made the structure less stable compared to the original d2 tetraloop with insertion (314-d2-tetraloop) but still more stable than the perfect d2 tetraloop with no stem defects (cutd2-tetraloop).

### *3.3.4.2 RMSD Analysis of de316-de318-314-d2-tetraloop*

All atoms RMSD values of de316-de318-314-d2-tetraloop were calculated and plotted in Figure 3.27. All atoms RMSDs showed that the structure of de316-de318-314-d2-tetraloop experienced similar changes to all atoms RMSDs of de316-314-d2-tetraloop. De316-de318-314-d2-tetraloop appeared to deviate slightly from its starting point soon after the simulation started and a little more after 4ns of simulation. The loop region and the stem region RMSD values of de316-de318-314-d2-tetraloop were calculated and plotted on the same graph for comparison (Figure 3.28).

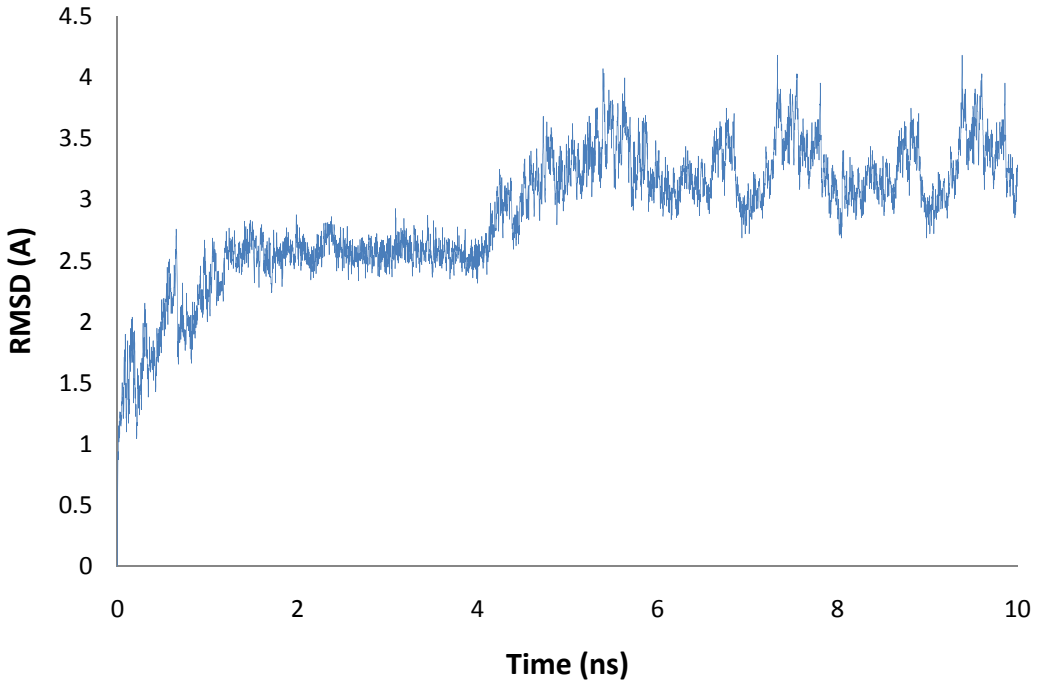



Figure 3.27: All atoms RMSD calculations of de316-de318-314-d2-tetraloop over 10ns of simulation time.

Figure 3.28 showed that it was the stem that experienced more changes after 4ns of simulation, which was averaged around 1.2 before 4ns and jumped to an average of 2.5 and fluctuated a lot. The loop of de316-de318-314-d2-tetraloop, on the other hand, had its RMSD values increased to over 3 Å soon after the simulation started, which could be explained by the missing loop-insertion hydrogen bonding observed in the simulation of de316-314-d2-tetraloop. Overall, the RMSD calculations of de316-de318-314-d2-tetraloop over 10ns of simulation were not much different from those of de316-314-d2-tetraloop. However compared to RMSDs of cutd2-tetraloop in Chapter II, it still remained a relatively stable stem-loop structure.

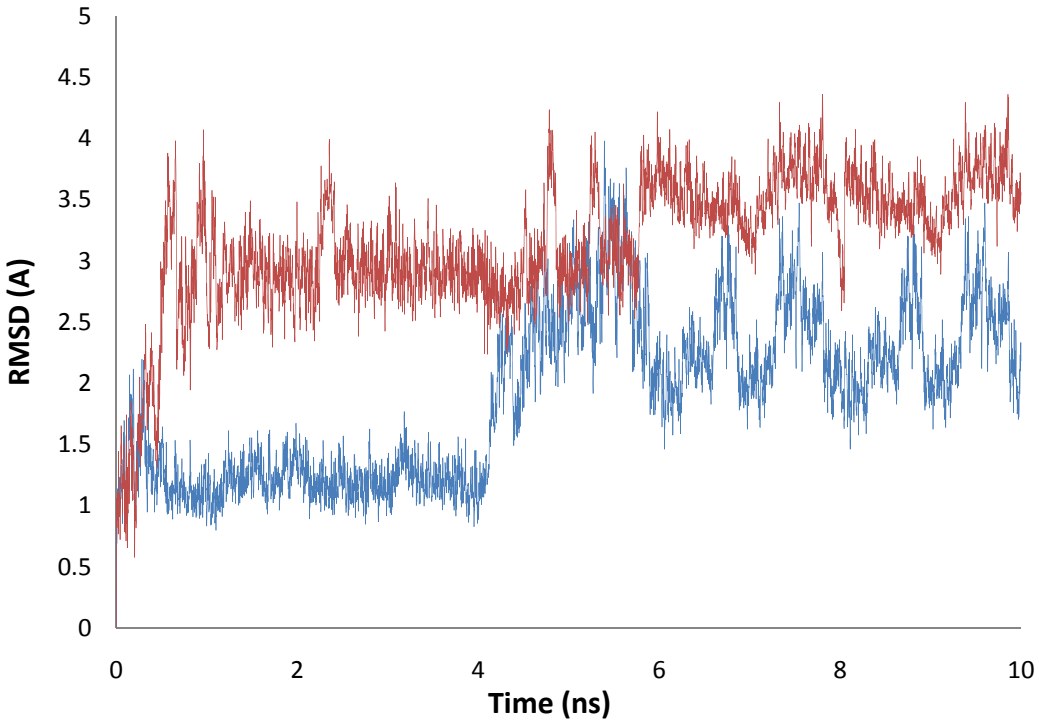



Figure 3.28: Loop (red) and stem (blue) region RMSD calculations of de316-de318-314-d2-tetraloop over 10ns of simulation time.

*3.3.4.3 Hydrogen bonding analysis*

The hydrogen bond interactions between the insertion residue and the loop observed previously in the simulation of de316-314-d2-tetraloop, such as 316A(N3) – 318C(H42), 316A(N3) – 318C(H41), and 316A(H21') – 318C(N4) were lost in the simulation of de316-de318-314-d2-tetraloop (Figure 3.29). The average distance between N3 on residue 316A and H41 on residue 318C, H21' on residue 316A and N4 on residue 318C, and N3 on residue 316A and H42 on residue 318C is 7.55 Å, 6.38 Å and 5.90 Å respectively. All three pairs of atoms fell way out of the distance cutoff of forming hydrogen bond interactions.

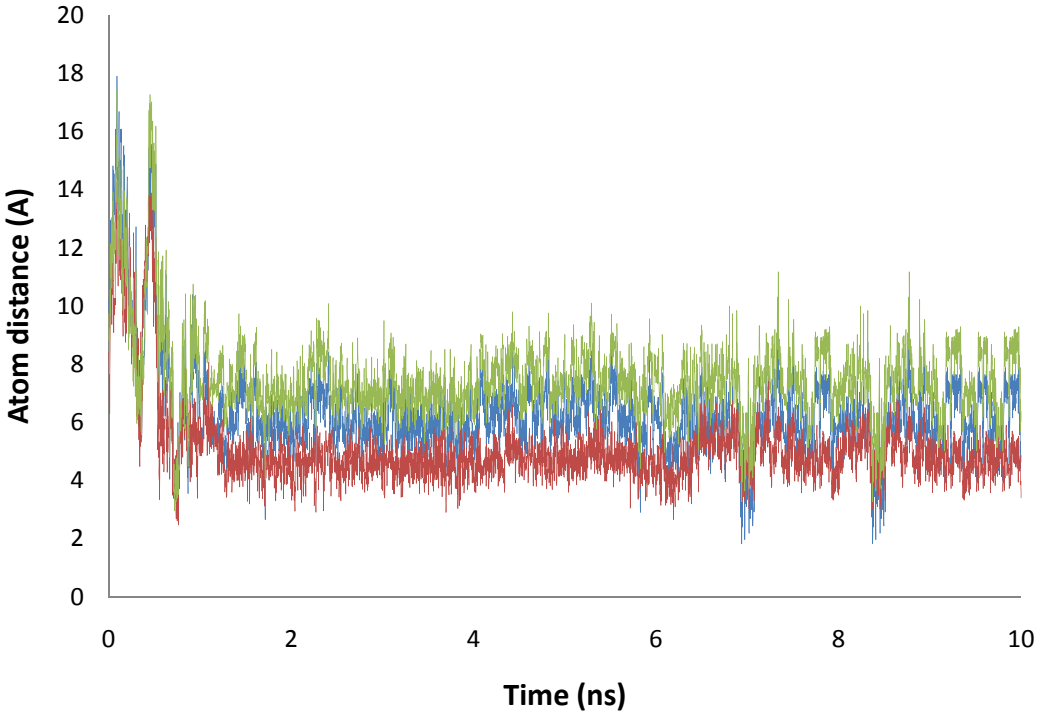



Figure 3.29: Distance calculated for atoms in the loop – insertion hydrogen bonding previously observed in the simulation of de316-314-d2-tetraloop, distance between N3 on residue 316A and H41 on residue 318C in blue, H21' on residue 316A and N4 on residue 318C in red, and N3 on residue 316A and H42 on residue 318C in green.

Again the hydrogen bonding calculation of de316-de318-314-d2-tetraloop showed that the intra-loop hydrogen bond interaction between residue j-1 and j+1 was lost soon after the simulation started (Figure 3.30). This could be explained by the missing hydrogen bond interactions between the loop and the insertion residue 318C and was consistent with the observed correlation of loop-insertion hydrogen bonding and intra-loop hydrogen bonding. The intra-loop hydrogen bonding calculation was also consistent with the loop region RMSD calculation. Both showed that the loop structure experienced considerable changes soon after the simulation started.

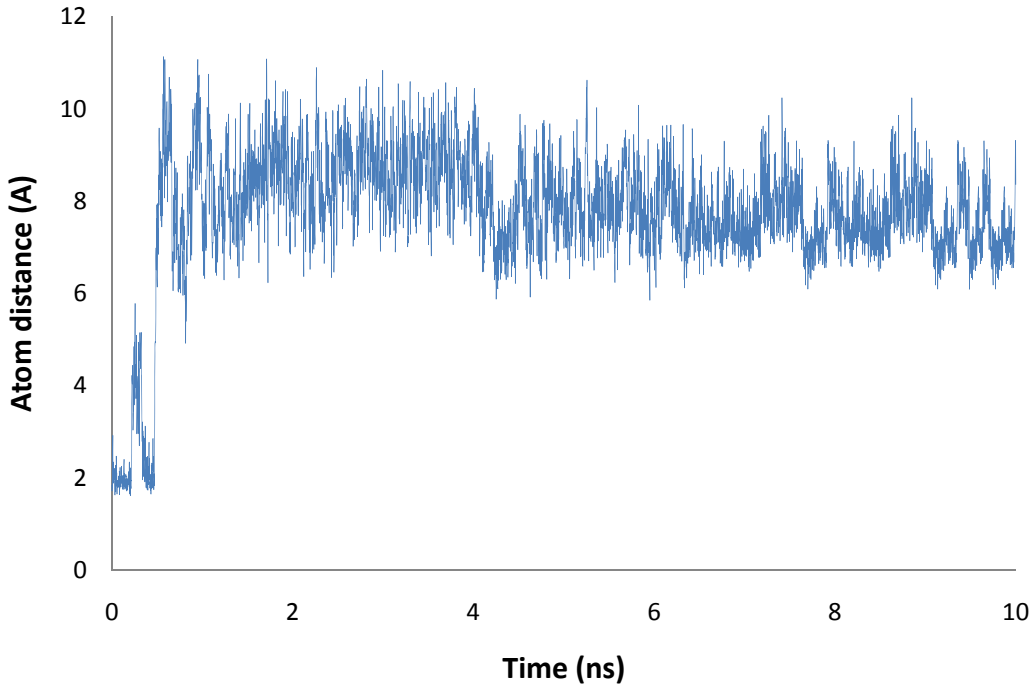



Figure 3.30: Intra-loop hydrogen bonding calculation between residue j-1 and residue j+1.

In sum, simulations of de316-de318-314-d2-tetraloop showed considerable changes in the loop region since the beginning of the simulation and minor changes in the stem region after 4ns of simulation. The loop-insertion hydrogen bonding interactions observed at the early stage of de316-314-d2-tetraloop simulation were absent in the simulation of de316-de318-314-d2-tetraloop, which could explain the deformed loop right after the simulation started.

Compared to simulations of cutd2-tetraloop conducted in Chapter II where the insertion residue 318C was completely removed from its stem, the structure of de316-de318-314-d2-tetraloop remained relatively stable. Therefore other than the enthalpy decrease due to both the loop-insertion and the stem-insertion hydrogen bonding interactions, conformational relaxation of the stem by the insertion residue is also a contributing factor for the stability of 314-d2-tetraloop.

## *3.4 Conclusions*

We validated our hypothesis II that the insertion residue 318C is the stabilizing factor of 314-d2-tetraloop, or stem defects are the stabilizing factors of d2 tetraloops in general. In the case of 314-d2-tetraloop, the insertion residue stabilizes the structure by forming hydrogen bonding with both the loop and the stem, which contributes to stability by decreasing system enthalpy. However our original reasoning on the stability contribution of these defects by conformationally relaxing the stem is also valid. By comparing the simulation of both de316-314-d2-tetraloop and de316-de318-314-d2-tetraloop, where only insertion hydrogen bonding interactions were removed, with the simulation of cutd2-tetraloop where the insertion residue was completely removed, we can see that cutd2-tetraloop could not remain stable at all. Therefore conformational relaxation of stem plays additional stabilizing role.

## *3.5 References*


1. Hsiao, C., Mohan, S., Hershkovitz, E., Tannenbaum, A., & Williams, L. D. (2006). Single nucleotide RNA choreography. *Nucl. Acids Res*. 34 (5), 1481–1491.
2. Klein,D.J., Schmeing,T.M., Moore,P.B. and Steitz,T.A. (2001) The kink-turn: a new RNA secondary structure motif. EMBO J., 20, 4214–4221.
3. Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 s resolution. Science, 289, 905–920.

4. Lee,J.C., Cannone,J.J. and Gutell,R.R. (2003) The lonepair triloop: a new motif in RNA structure. J. Mol. Biol., 325, 65–83.
5. Jorgensen, W.L., et al., Comparison of simple potential functions for simulating liquid water. Journal of Chemical Physics, 1983. 79: p. 926-935.
6. Cornell, W. D., Cieplak P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., & Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117 (19): 5179-5197, 1995.
7. Hershkovitz, E., Tannenbaum, E., Howerton, S. B., Sheth, A., Tannenbaum, A. and Williams, L. D. (2003). Automated identification of RNA conformational motifs: theory and application to the HM LSU 23S rRNA. *Nucleic Acids Research*, 31(21), 6249-6257.

# CHAPTER IV

# FOLDING PATHWAY OF TETRALOOP

This chapter presents the study of folding pathway of RNA standard tetraloops through molecular dynamics simulation using the method of thermal unfolding.

## *4.1 Introduction*

### 4.1.1 Background

*4.1.1.1 The Hierarchical Folding of RNA Structure*

The search for rules in protein folding to predict the three-dimensional structures of proteins from their primary sequences has attracted scientists for almost five decades. Structural prediction in protein folding is difficult because it takes 20 different amino acid residues to build a protein molecule. Thus, the number of distinct interactions among these residues is large, depending not only on the nature of the residues (hydrophobic, hydrophilic, polar, etc.), but also on each residue's detailed structure. Another difficulty is that the existence of the various secondary structural elements (α-helices, parallel and anti-parallel β-sheets, β-turns, random coils, etc.) is contextual, i.e. these elements form and are stable in the context of the rest of the protein, but may not form when they are isolated in solution. The formation of secondary structure depends on the nature of the tertiary folding contacts, and vice versa. This implies that the contributions of secondary and tertiary interactions to the energetic stability of a protein are, in principle, not separable. Therefore, no simple rules may exist to predict the three-dimensional structure of the protein from its

sequence.

RNA folding is simpler compared to protein folding. Only four nucleotides, each made of a ribose, a base, and a phosphate, are used as building blocks of the structure. Each ribose is a five-membered ring with only two main conformations. The structures of four bases are very similar to one another. The interactions of bases are either through hydrogen bonding or base stacking. The phosphate groups with their long range electrostatic interactions provide theoretical difficulties not present in protein folding. However, they actually simplify the RNA folding problem since electrostatic interactions are well understood theoretically.

The differences between protein and RNA folding can be viewed in terms of the distinct ways in which the information contained in the sequence of these two molecular species controls the secondary and tertiary structures. The information in the amino acid sequence of a protein is believed to branch into secondary and tertiary structural elements as well as flow between these elements in both directions. The information in the sequence of RNA, in contrast, flows linearly, and largely in one direction, first to the secondary and then to the tertiary structure. Therefore, an RNA molecule can be thought of as possessing a hierarchical structure in which the primary sequence determines the secondary structure which, in turn, determines its tertiary folding, whose formation alters only minimally the secondary structure [1].

#### *4.1.1.2 Formation of RNA Hairpin Secondary Structure*

The hairpin structure is the most common of the secondary structural elements in RNA. In its simplest form, the hairpin consists of a stem and a loop (Figure 4.1). The stem consists primarily of Watson-Crick base pairs formed between the two antiparallel stretches of RNA. The region of unpaired nucleotides at the apex of the structure is the hairpin loop. The most obvious function of the loop is to reverse the directionality of the backbone so that it can afford the antiparallel strands of the stem. Because of steric

repulsion, it takes a minimum of three nucleotides to make a turn in RNA hairpin, although loops of four nucleotides (tetraloops) are much more common.

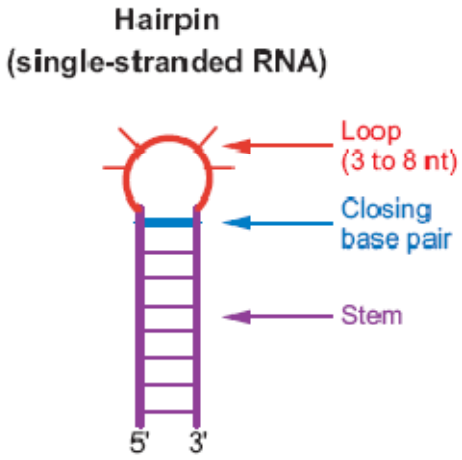



Figure 4.1: Stem loop sketches of hairpin structure [2].

Tinoco and Bustamante suggested that the initiating of RNA hairpin folding is the random collision of complementary base pairs [3]. The formation of loop causes great entropy loss, with a hairpin loop with four nucleotides having an increase in free energy of 4.5 kcal/mole. Therefore base interactions, such as hydrogen bonding or base stacking, need to be formed to compensate for the entropy loss so it can close a hairpin loop. The rate of hairpin loop formation thus depends on the frequency of collision between two complementary bases, which depends on the effective concentration of one base relative to another, i.e. on the effective relative volume that is accessible to two bases. The effective relative volume depends on the number of nucleotides separating the two bases. Therefore the kinetics of folding favors the formation of small hairpin loop, such as tetraloop.

### 4.1.2 Motivation

*4.1.1.1 Functions of RNA Hairpins in Biology*

The understanding and appreciation of RNA biology have undergone a revolution over the past few decades. It is now recognized that RNA is much more than a passive intermediary between DNA and proteins. RNA can both store genetic information and perform varied functions. These discoveries have led to the notion that RNA, or an RNA-like polymer, played a central role in the emergence of life on Earth – the so-called RNA world hypothesis [4–6].

The motivation behind studying RNA folding is to elucidate its varied functions in biology. It is now known that RNA plays important biological roles in all levels of its folding hierarchy, from unstructured, to base paired (hairpin), and finally to globular tertiary structures, which typically contain multiple hairpins [2].

RNA can function in biology when it has little or no self-structure. For example, noncoding RNAs, such as mature miRNAs, control gene expression by base pairing to complementary coding and noncoding regions on mRNA [7, 8]. In addition to pairing with other RNAs, unstructured RNA molecules can bind proteins in a sequence-specific fashion to effect biological functions such as splicing, development, and gene regulation [9–13].

Although RNA can be found in nature without significant secondary structure, most RNA strands spontaneously fold into secondary structures, for example hairpins, that offer additional layers of biological functions such as controlling transcription termination and gene expression [11]. RNA hairpins are especially important in controlling gene expression. Fundamentally, this is because RNA can switch between two conformations, one of which allows gene expression and one of which blocks it.

When adopting tertiary structures in which secondary structures interact with

each other through space to give compact, solvent-excluded structures with binding pockets and active sites, RNA molecules can perform the most complex cellular functions, such as small molecule binding, protein synthesis, splicing, and tRNA maturation [1, 14 – 16].

In general, RNA molecules fold in a hierarchical fashion which will be discussed in the next section. Therefore it is very important to study its folding.

### 4.1.3. Objectives

This chapter focuses on studying the folding of a standard tetraloop in hope of revealing the general folding mechanism of RNA hairpin structure. Another objective is to find any intermediate in the folding pathway. Studies on tetraloop folding have shown that three states were observed during the folding: unfolded state, globular intermediate state, and folded state [17, 18]. In Chapter III, we discussed high abundance of d2 tetraloops in the HM 23S rRNA database and revealed some of the factors keeping this structure stable. We believe another explanation for such high abundance of d2 tetraloops is that it is a kinetically trapped intermediate during folding of a standard tetraloop.

*Hypothesis III. A d2 tetraloop is a kinetically trapped intermediate in the folding of a standard tetraloop.*

To test hypothesis III, we simulated the folding of a sample structure of the standard tetraloop from the HM 23S rRNA database, 2249-standard tetraloop.

## *4.2 Method*

### 4.2.1 Secondary Structure of 2249-Standard Tetraloop

Our study of the folding mechanism of standard tetraloops was conducted on a sample structure of standard tetraloops, 2249-standard tetraloop, from HM 23S ribosomal RNA database. This 2249-standard tetraloop has a sequence of UCG(GGGA)CGA (Figure 4.2). Again the number 2249 represents the residue number of the first nucleotide in the loop region.

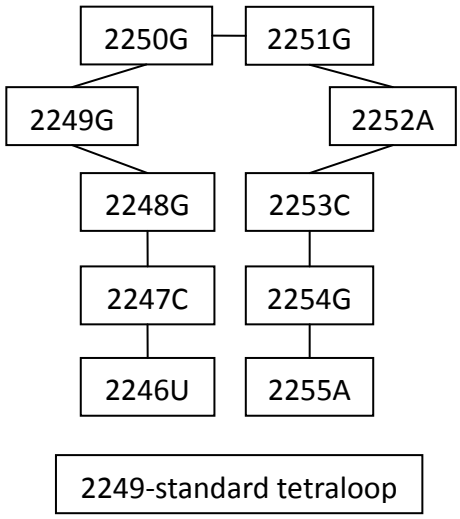



Figure 4.2: Secondary structure of 2249-standard tetraloop

This 2249-standard tetraloop (GGGA) belongs to the class of GNRA tetraloops, which has extraordinary thermodynamic stabilities and has a lot of very important biological functions. Therefore simulations on the folding of this chosen

sample structure can give meaningful insights into the folding mechanism of the RNA tetraloop in general.

### 4.2.2 Thermal Unfolding Simulation

Molecular dynamics simulation methods have provided key insights into the general nature of biomolecule motion, including protein and nucleic acids, and aspects of motion linked to the function of those molecules in their native state. The molecular dynamics methods using detailed atomic models for the molecules and solvent have been applied extensively in order to explore the process of biomolecules folding and unfolding.

Unfortunately molecular dynamic simulations of RNA folding can be extremely expensive in terms of computational time. Meanwhile micro- to millisecond timescales are often necessary to simulate folding events of even the smallest RNA secondary structure units. High temperature can be used as a means of accelerating the process. Raising the temperature has a small effect on velocity (atoms at 498K move approximately 30% faster than at 298K) but activated processes involving traversing energy barriers will be greatly accelerated. The acceleration depends on the enthalpy of activation and can give increase in speed of $10^3$ ($\Delta H^{\ddagger}$ = 10 kcal/mol) to $10^9$ ($\Delta H^{\ddagger}$ = 32 kcal/mol) [21].

Therefore, in this study, high temperature induced unfolding molecular dynamics simulations of 2249-standard tetraloop were carried out to provide potential insights into folding pathways between the native folded and denatured unfolded states of the system and to assess folding dynamics within the tetraloop structure. The reverse of high temperature simulation trajectory of this 2249-standard tetraloop should reveal an approximation of its folding pathway and any intermediates in the folding pathway.

Thermal unfolding simulation has served as a powerful tool for investigating dynamics involving large conformational changes on computationally tractable timescales. It has been previously suggested that the unfolding process for proteins can in general reflect the main attributes of the folding event [19 – 23], and it has been shown that construction of folding pathways, including transition state ensembles, is possible using high temperature induced unfolding [20]. A report of the direct, atomistic folding of a small protein *in silico* [24] has shown that previously published thermal unfolding trajectories [20] served as a good tool for predicting the folding rate and mechanism.

### 4.2.3 Simulation Protocols

NPT MD simulation in explicit water (TIP3P [25]) was performed following the same simulation protocol used in Chapter II and Chapter III. Again AMBER94 [26] was the force field used.

Test simulations were conducted at four different temperatures, 410K, 510K, 610K and 710K for 10 nanoseconds in order to determine the appropriate temperature and simulation length. The simulation conducted at 410K showed that the structure of 2249-standard tetraloop did not unfold completely. The simulations conducted at 610K and 710K showed that the structure reached its denatured state with 100ps and no intermediates were observed. Trajectory of the simulation conducted at 510K, on the other hand, best represented an ensemble of unfolding events. It appeared that 510K was the lowest temperature that caused the RNA to completely unfold, which was consistent with most of the thermal unfolding experiments from the literature.

In addition, testing simulations showed that at 510K the RNA tetraloop molecule reached its completely unfolded state during the first 2 ns of simulation. Therefore all high temperature simulations in this study were run at constant temperature 510K for 2 ns.

## *4.3 Results and Discussions*

High temperature MD simulations at 510K following the same simulation protocol described earlier were repeated for four runs in this study. Trial 1 and 4 revealed the same unfolding event, here called unfolding pathway 1. Trial 2 revealed a different unfolding event, here called folding pathway 2. Both folding mechanisms are discussed here.

### 4.3.1 Unfolding Pathway 1

Two runs of high temperature MD simulations at 510K, trial 1 and trial 4, showed similar trajectories, here called unfolding pathway 1. The observed unfolding events in unfolding pathway 1 displayed three-state folding behavior, including a folded (native) state, an intermediated state, and an unfolded state. RMSDs analysis and hydrogen bonding calculations were conducted to exhibit how 2240-standard tetraloop unfolded itself during the simulation.

*4.3.1.1 RMSD Analysis of Unfolding Pathway 1*

All atoms RMSD values of 2249-standard tetralop from unfolding simulation 1 were calculated and plotted in Figure 4.3. The all atoms RMSDs calculation in unfolding pathway 1 clearly displayed three-state folding behavior of 2249-standard tetraloop: a native (or folded) state, an intermediate state, and an unfolded state. The structure of 2249-standard tetraloop started to unravel very quickly during the first 200ps of high temperature simulation with its RMSDs increasing to 5 Å. Then the structure reached its intermediate state with RMSDs leveling off and staying at around 2 – 3 Å until 1ns. After 1ns of the simulation the structure of 2249-standard tetraloop reached its unfolded state with continuously increasing RMSDs for the rest of the simulation.

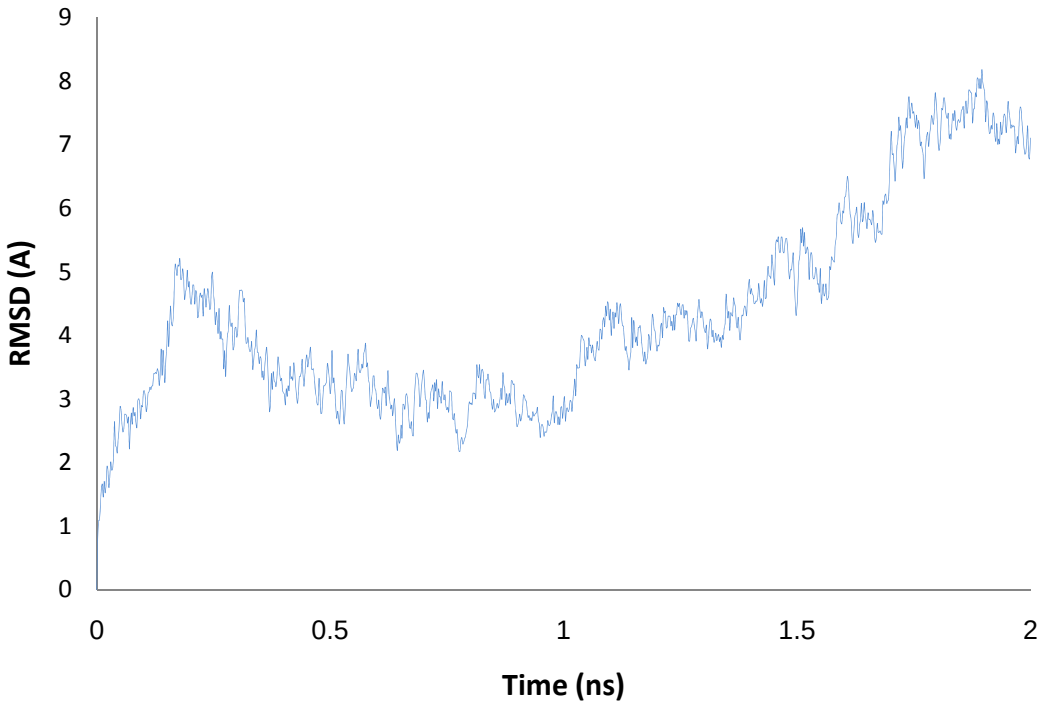



Figure 4.3: All atoms RMSD calculations for high temperature unfolding simulation 1.

RMSD values of the loop and the stem region during the high temperature simulation were also calculated and plotted on the same graph for stem-loop comparison (Figure 4.4). Generally both the loop and the stem RMSD calculations showed similar patterns of the structural unfolding for all atoms RMSD calculations (Figure 4.3). However, the stem appeared to be experiencing more atomistic changes than the loop during the unfolding, which leads us to believe that the stem is less stable than the loop when the structure unfolds itself and is the last to form during the folding of tetraloops.

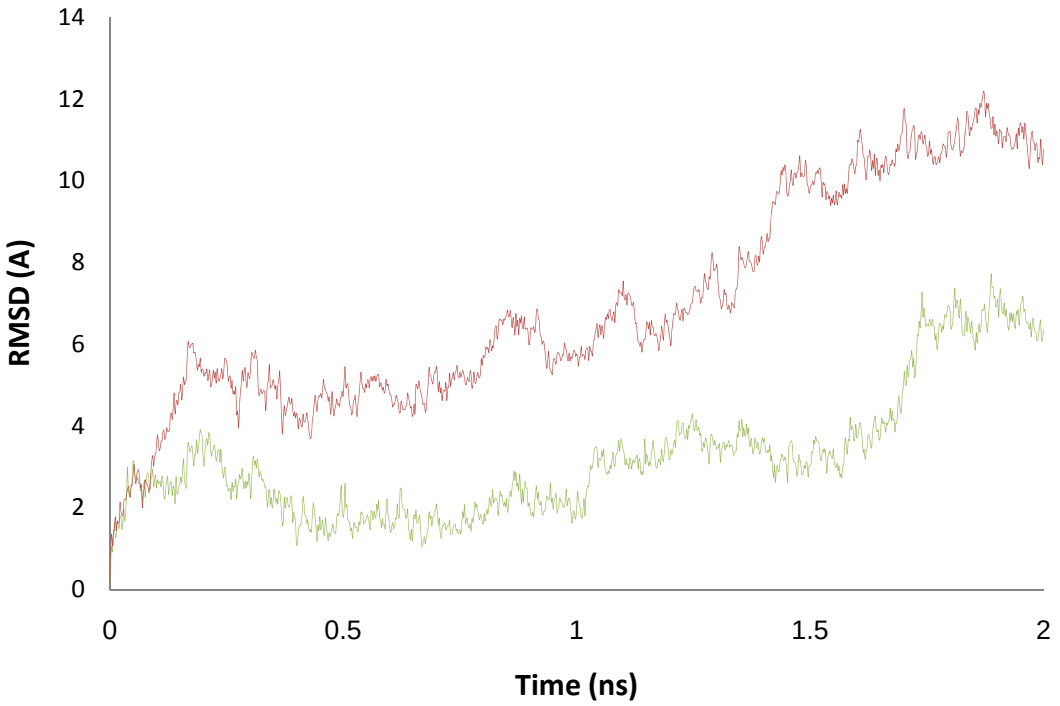



Figure 4.4: RMSD calculations of the loop region (red) and the stem (green) over 2ns of simulation time for high temperature unfolding simulation 1.

RMSDs of each base pair in the stem of 2249-standard tetraloop were calculated and plotted in Figure 4.5 and were consistent with the stem RMSD calculation in Figure 4.4. The atom positions of the terminus base pair started to deviate from their starting point appreciably soon after the simulation started, with terminus base pair RMSD increased to over 4 Å after 200ps. The central and the closing base pair RMSD remained relatively stable during the first 1ns and then jumped high after 1ns. Hydrogen bonding interactions between each base pair in the stem were calculated in the next section.

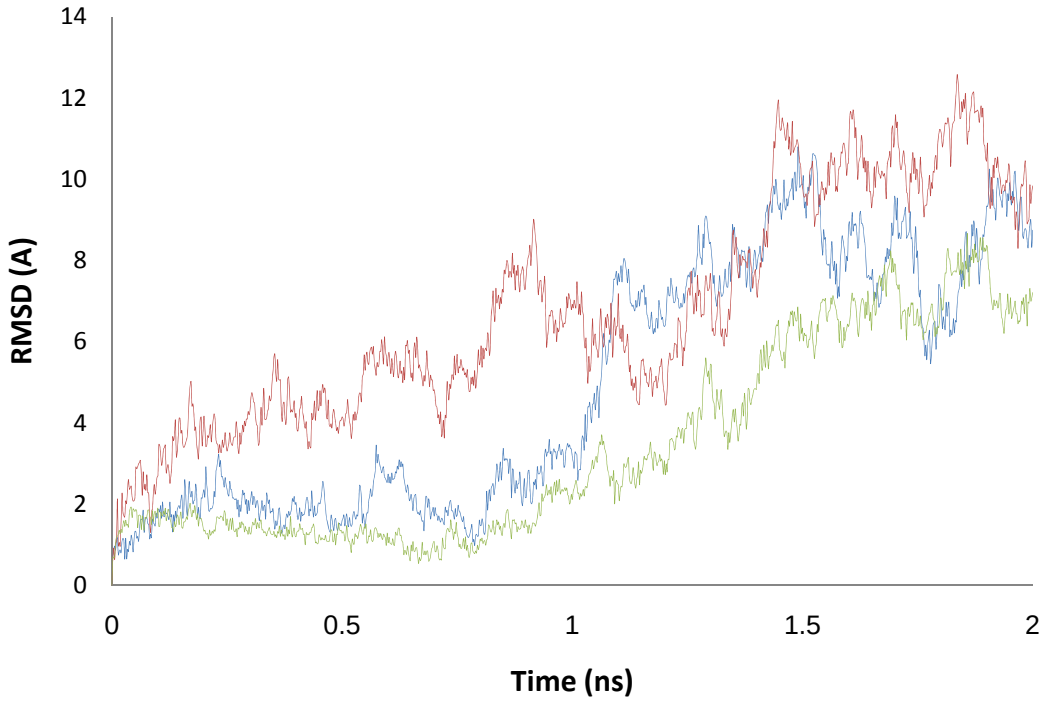



Figure 4.5: RMSD calculations for three base pair in the stem over 2ns of simulation time for unfolding simulation 1. The terminus base pair (2246U-2255A) in red, the central base pair (2247C-2254G) in blue, and the closing base pair (2248G-2253C) pair in green.

In summary, RMSD analysis of 2249-standard tetraloop high temperature simulation trajectory displayed a three state folding behavior: a native (folded) state, an intermediate state, and an unfolded state. The unfolding trajectory of 2249-standard tetraloop showed a less stable stem structure relative to the loop. The relative stability of the loop compared to the stem and the order of base pair opening up in the stem (the terminus base pair followed by the central base pair and then the closing base pair) suggested that the loop probably forms first followed by the stem during the folding of tetraloops.

#### *4.3.1.2 Hydrogen Bonding Interactions*

Hydrogen bonding interactions of the terminus base pair (2246U-2255A), the central base pair (2247C-2254G) and the closing base pair (2248G-2253C) were calculated and plotted in Figure 4.6, Figure 4.7 and Figure 4.8. The terminus base pair opened immediately after the simulation started. The central base pair held much longer than the terminus base pair during the simulation and started to open up at 800ps. The closing base pair seemed to be the most stable base pair and held the longest among all three base pairs in the stem. It started to open up after 1ns of simulation.

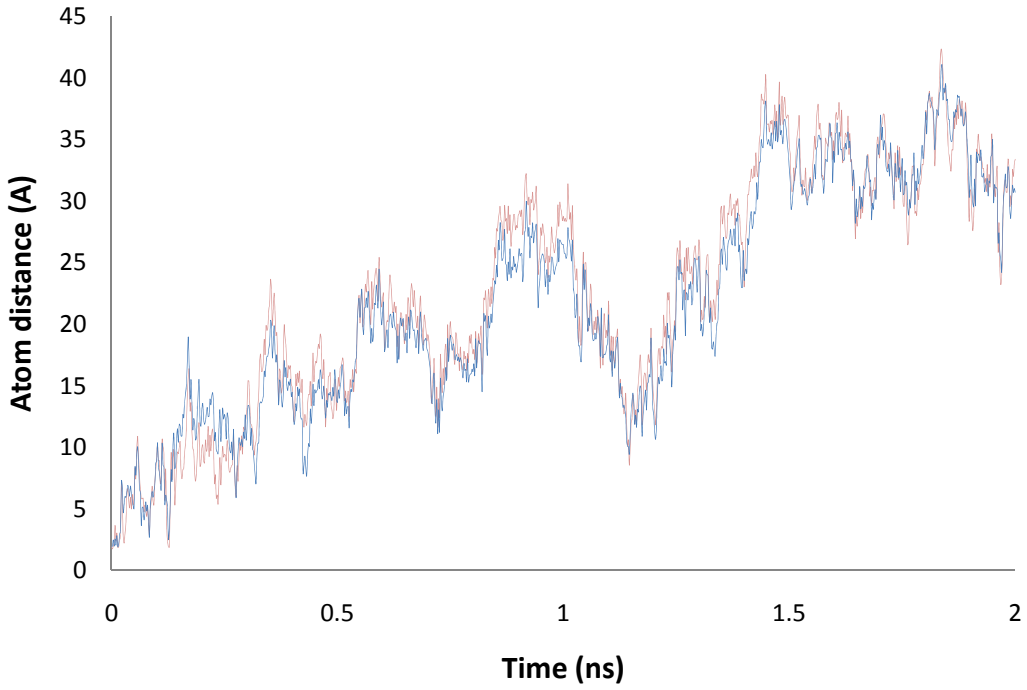



Figure 4.6: Hydrogen bonding interactions between the terminus base pair in the stem of unfolding simulation 1. Distance of O4 on 2246U and H61 on 2255A in pink, distance of H3 on 2246U and N1 on 2255A in blue.

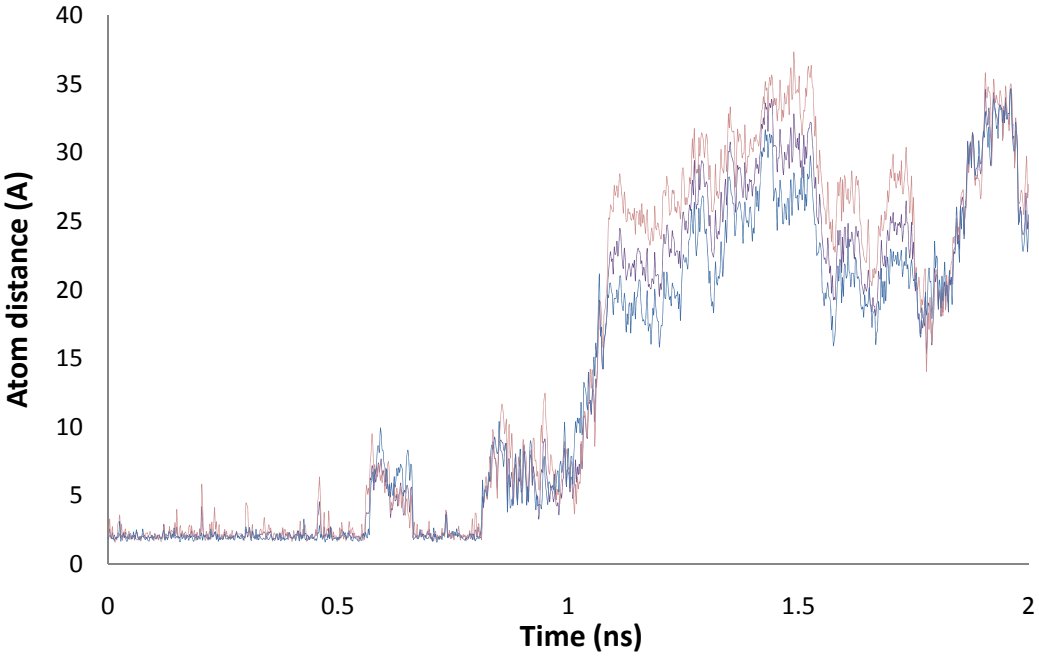



Figure 4.7: Hydrogen bonding interactions between the central base pair in the stem. Distance of N3 on 2247C and H1 on 2254G in purple, distance of H41 on 2247C and O6 on 2254G in pink, and distance of O2 on 2247C and H2 on 2254G in blue.

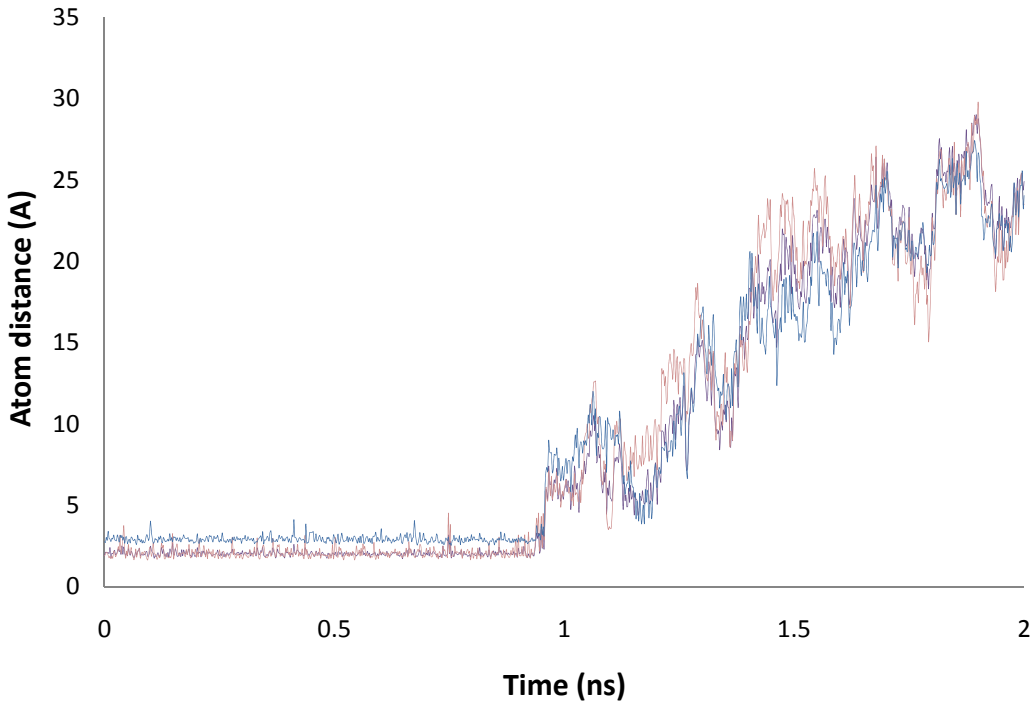



Figure 4.8: Hydrogen bonding interactions between the closing base pair in the stem. Distance of H1 on 2248G and N3 on 2253C in purple, distance of O6 on 2248G and H41 on 2253C in pink, and distance of H21 on 2248G and O2 on 2253C in blue.

Hydrogen bonding interaction calculations of each base pair in the stem of 2249-standard tetraloop were consistent with the RMSD analysis. The structure of 2249-standard tetraloop was completely unfolded in the stem region after 1ns of high temperature simulation. Consistent with the calculations of stem base pairs RMSDs, hydrogen bonding of the terminus base pair was first broken, suggesting that terminus bases did not pair in the intermediary state. The central base pair was still held in the intermediary state. The closing base pair was conserved for the longest time among three base pairs. The closing base pair opening up indicated that the structure of 2249-standard tetraloop reached the unfolded state.

The intra-loop hydrogen bonding interactions, the one between residue j-1 and residue j+1 and the one between residue j-1 and residue j+2, were calculated in Figure 4.9 and Figure 4.10. Unlike what we expected, the intra-loop hydrogen bonding interaction between residue j-1 and j+1, which is believed to be a characterizing interaction for tetraloops structure, were broken soon after the simulation started and did not exist in the intermediate state observed in unfolding pathway 1. If d2 tetraloops with some kind of defects in the stem, which are found abundantly in the database, was indeed an intermediate state during the folding pathway of a standard tetraloop as we proposed, then the intra-loop hydrogen bonding interaction between j-1 and j+1 would have been conserved during the intermediate state, at least before the structure went completely unfolded.

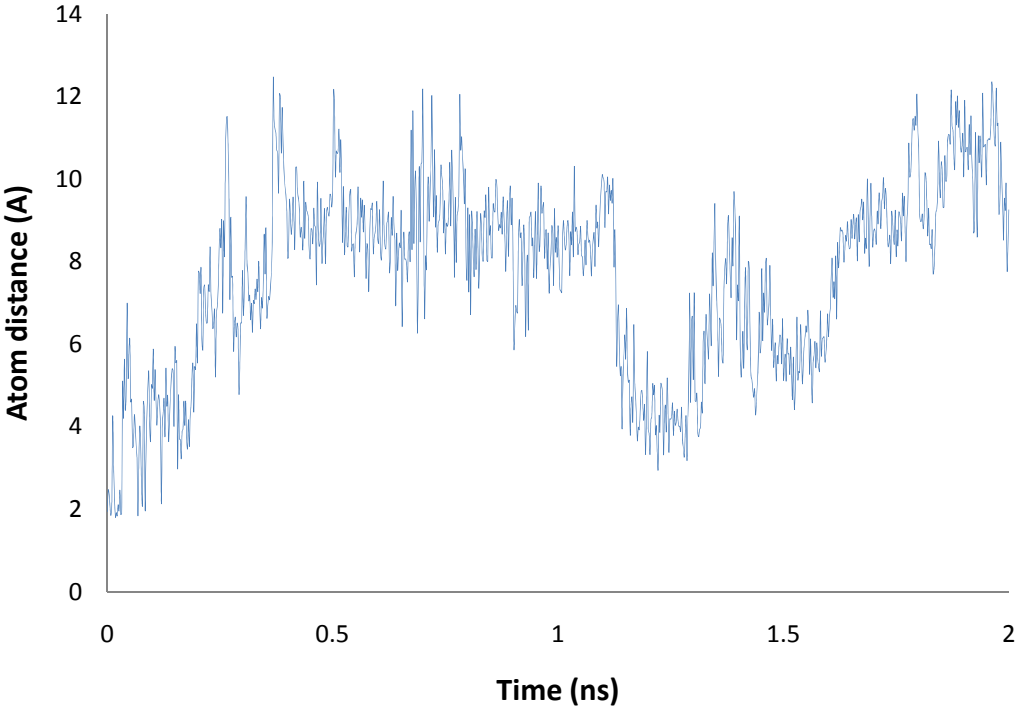



Figure 4.9: Intra-loop hydrogen bonding interactions between 2249G (j-1) and 2251G (j+1) from unfolding simulation 1.

The other intra-loop hydrogen bonding interaction between residue j-1 and j+2 was conserved relatively well for the first 1.1ns of the simulation but was no longer conserved after that. Recall that the hydrogen bonding interactions between the closing base pair were broken at approximately 1ns of the simulation, indicating the opening up of the closing base pair. It was after the lost hydrogen bonding interactions between the closing base pair that this j-1 – j+2 intra-loop hydrogen bonding interaction also started to break up. Therefore we can say that closing base pair appears to be critical to hold the loop structure even though this loop is no longer considered a tetraloop without the j-1 – j+2 intra-loop hydrogen bonding interaction in place. It also showed that the opening up of closing base pair caused the opening up of the loop and led 2249-standard tetraloop to reach the completely unfolded structure.

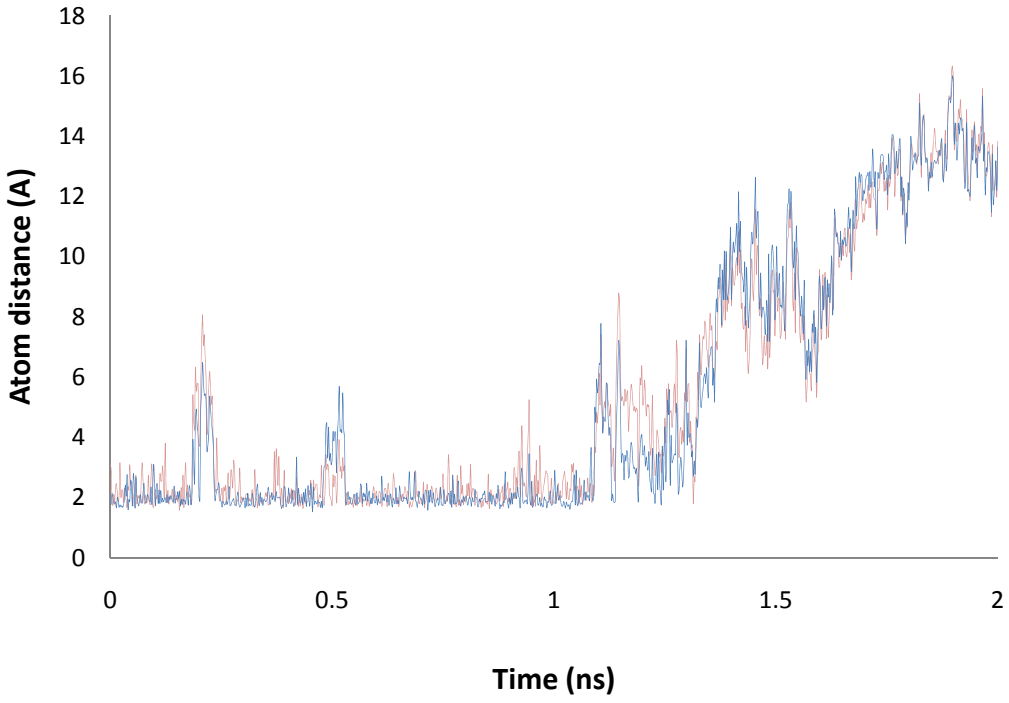



Figure 4.10: Intra-loop hydrogen bonding interactions between 2249G (j-1) and 2252A (j+2) from unfolding simulation 1, atom distance between H1 of 2249G and O2P of 2252A in pink and atom distance between H21 of 2249G and O2P of 2252A in blue.

In summary, RMSD analysis and hydrogen bonding calculations of 2249-standard tetraloop high temperature simulation 1 supported the theory of loop forms first followed by the stem during the folding of tetraloops. The overall unfolding trajectory is described in detail in the next section.

#### *4.3.1.3 High Temperature Unfolding Trajectory 1*

Two nanosecond of high temperature MD simulation trajectory was shown in Figure 4.11, revealing an observed unfolding pathway. Twenty one snapshots were taken from the starting point to the end point of 2ns with a 100ps increment. Simulation time corresponding to when the snapshot was taken was noted below each snapshot. The phosphate-ribose backbones of the stem region were displayed as blue sticks for clarity. The phosphate-ribose backbones of the loop region were colored according to their residue ID, 2249G (j-1) in pink, 2250G (j) in blue, 2251G (j+1) in purple, and 2252A (j+2) in green.

The unfolding pathway 1 showed the loss of the terminus base pair (2246U – 2255A) occurred first just after the simulation started, accompanied by the loss of intra-loop hydrogen bonding interaction between 2249G (j-1) and 2249G (j+1). At this time the backbone of the loop became slightly flattened, however, the loop still remained because of the closing base pair and the other intra-loop hydrogen bonding interaction between 2249G (j-1) and 2252A (j+2). This state was identified as the intermediary state which lasted for approximately 800ps during the high temperature simulation. After 800ps the central base pair (2247C – 2254G) started to open up, followed by the loss of closing base pair (2248G – 2253C). Then the j-1 – j+2 intra-loop hydrogen bonding was lost as well as the loop structure. After approximately 1.2ns, the structure reached its completely unfolded state.

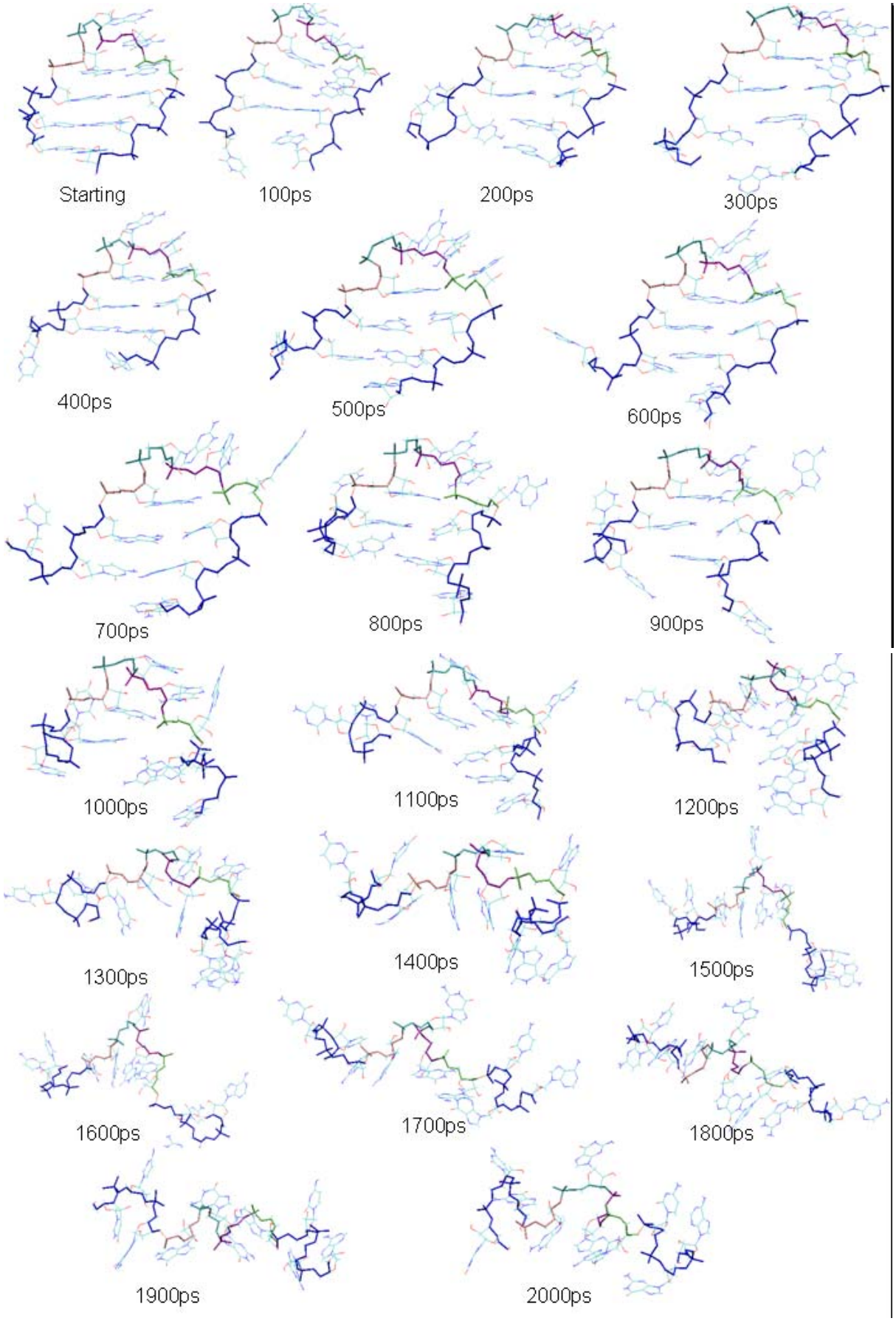



Figure 4.11: Two nanosecond high temperature (510K) simulation trajectory 1.

Reversing the high temperature unfolding trajectory 1 gives a relatively clear folding pathway of a standard tetraloop, 2249 standard tetraloop in this case. The unfolded RNA single strand forms a loop first. This is achieved by random collision of complementary base pair to form the closing base pair first. Then the stem starts to grow by forming base pairing and stacking. Then the loop starts to 'settle down' to the loop structure of a tetraloop by forming all the right intra-loop hydrogen bonding interactions and the stacking of bases on residue j, j+1, and j+2. Meanwhile the stem finishes its structure by forming the terminus base pair. At this point the folding of a standard tetraloop is complete.

### 4.3.2 Folding Pathway 2

The other run (trial 2) of high temperature MD simulations at 510K showed a different trajectory, here called unfolding pathway 2. The observed unfolding events also displayed three-state behavior, including a folded (native) state, an intermediate state, and an unfolded state. What is different this time is that the structure of this observed intermediate state is similar to that of the d2 tetraloop, which is what we initially proposed. However, this intermediate state remained for only a very short period of time (about 100ps) during the high temperature simulation. Afterwards the structure reached the unfolded state.

#### *4.3.2.1 RMSD Analysis of Unfolding Simulation 2*

All atoms RMSD values from 2249-standard tetraloop unfolding simulation 2 were calculated and plotted in Figure 4.12. The structure started to unravel very quickly

during the first 100ps with the RMSDs increased to approximately 4 Å. After 300ps the RMSD started to drop to less than 2 Å and reached its lowest point (1.56 Å) at 500ps, indicating that the structure of 2249-standard tetraloop reached its intermediate state. Afterwards the RMSDs went up continuously to over 5 Å. Therefore, the all atoms RMSDs of unfolding simulation 2 also showed a three state folding behavior: a native state, an intermediate state, and unfolded state.

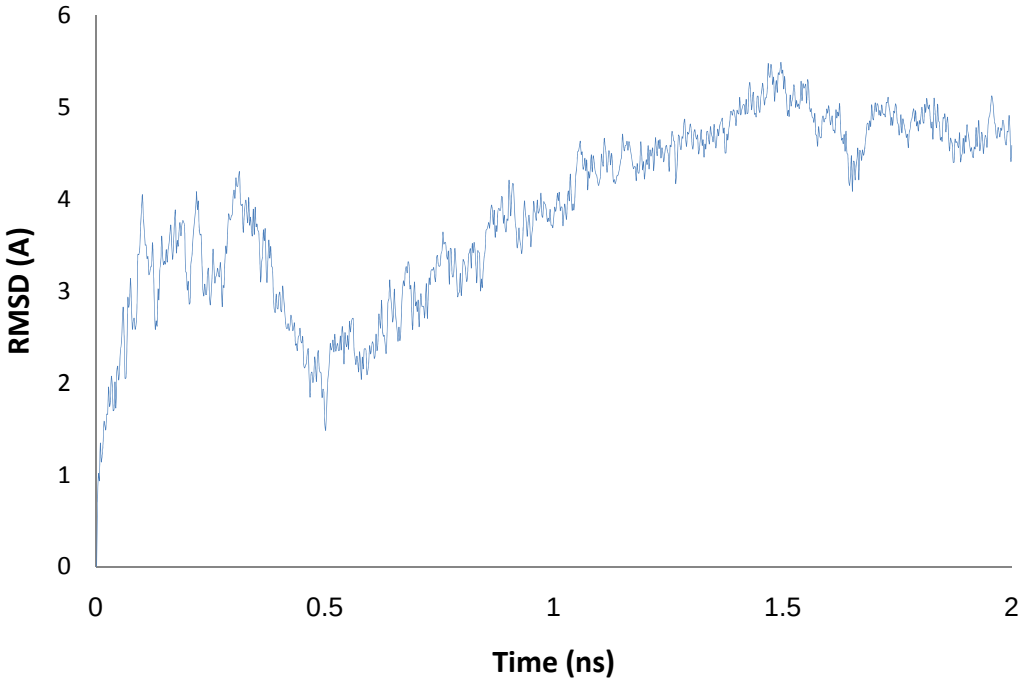



Figure 4.12: All atoms RMSD calculations from high temperature unfolding simulation 2.

In comparison with the intermediate state observed in unfolding simulation 1, this intermediate state in the unfolding simulation 2 did not last very long (for approximately 200ps from observation of all atoms RMSDs). The intermediate state from the high temperature unfolding simulation 1 lasted for a much longer time (for approximately 1ns).

The loop and the stem region RMSDs were calculated and plotted in the same graph (Figure 4.13). The pattern of stem RMSDs were very similar to that of all atoms RMSDs (Figure 4.12). The pattern of the loop RMSDs, on the contrary, is quite different. It had its first jump at approximately 200ps, at which point the RMSDs of the stem decreased, and the second jump after 500ps of simulation.

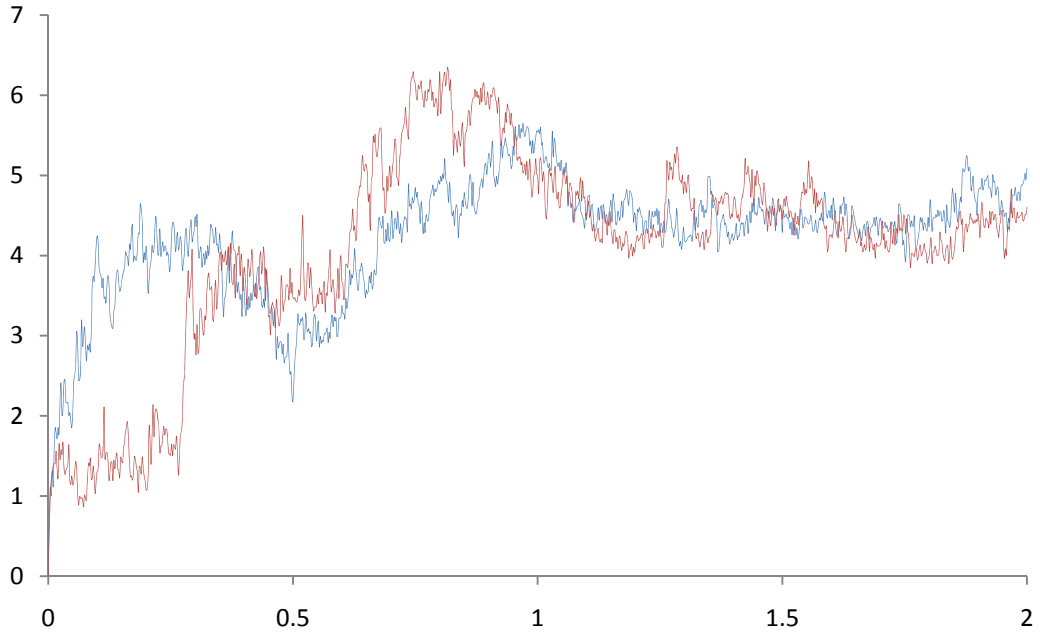



Figure 4.13: RMSD calculations of the loop region (red) and the stem (blue) over 2ns of simulation time for high temperature unfolding simulation 2.

The loop RMSDs calculation was consistent with the observation that the structure of the intermediate state resembled that of d2 tetraloops. After 200ps 2249-standard tetraloop reached its intermediate state and the structure of the loop changed appreciably (the first jump in loop RMSDs). To better understand the structural change of the loop during high temperature simulation RMSDs of residue

2249G (j-1), 2250G (j) and 2251G (j+1) were calculated in Figure 4.14. It showed that residue j-1, j and j+1 went back to its starting point in the intermediate state but not residue j+2 (see RMSDs of the whole loop in Figure 4.13). Therefore, the loop of the intermediate state is structurally the same as the loop of the d2 tetraloop.

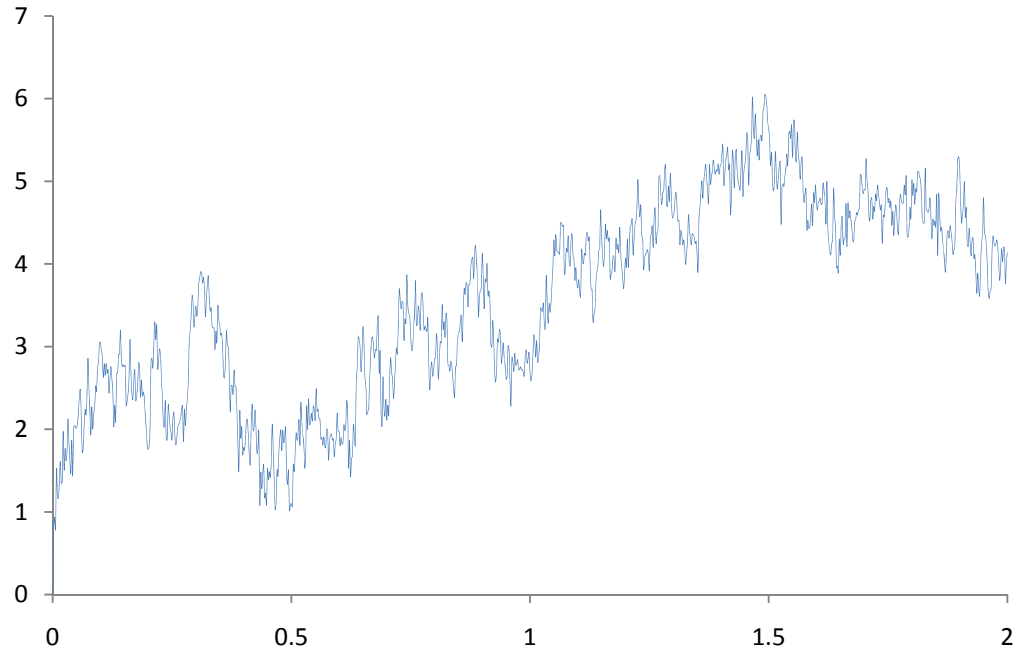



Figure 4.14: RMSD calculations of residue j-1, j, and j+1 in the loop region from unfolding simulation 2.

The RMSDs of each base pair in the stem were calculated and plotted in the same graph (Figure 4.15). The stem base pairs RMSDs calculations showed that the atom positions of both the terminus base pair and the central base pair changed appreciably soon after the simulation started. The atom positions of closing base pair remained relatively intact until 800ps, suggesting again that the closing base pair was the last base pair to open up during the unfolding of 2249-standard tetraloop. Hydrogen bonding interactions between all three base pairs in the stem were calculated in the next section.

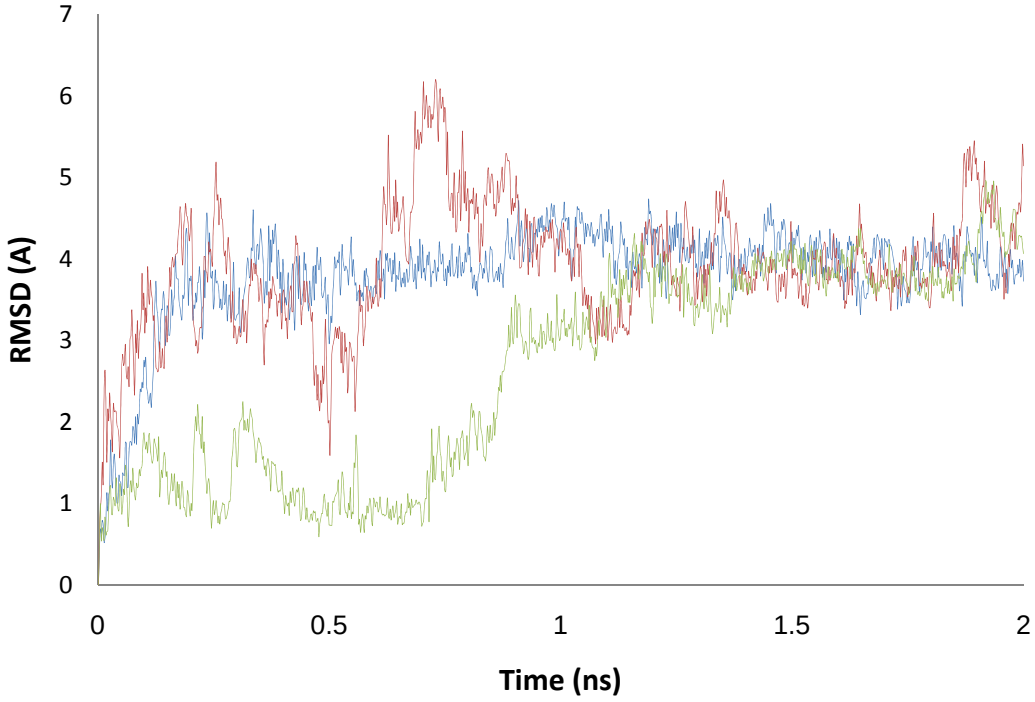



Figure 4.15: RMSD calculations for all three base pairs in the stem for unfolding simulation 2, closing base pair (2248G – 2253C) in green, central base pair (2247C – 2254G) in blue, and terminus base pair (2246U – 2255A) in red.

In summary, the RMSD analysis of high temperature simulation 2 exhibited a three-state unfolding behavior: a native state, an intermediate state and an unfolded state. The intermediate state observed in unfolding simulation 2 resembled the structure of the d2 tetraloop. RMSDs calculations of both the loop of 2249-standard tetraloop and residue j-1, j and j+1 in the loop of 2249-standard tetraloop further suggested that the intermediate state is a d2 tetraloop. Therefore, it validated our hypothesis III that a d2 tetraloop is a kinetically trapped intermediate in the folding of a standard tetraloop.

*4.3.2.2 Hydrogen Bonding Calculation of Unfolding Simulation 2*

Hydrogen bonding interactions in each stem base pair were calculated and showed consistency with the RMSDs analyses of stem base pairs. The terminus base pair (2246U-2255A) opened up immediately after the simulation started (Figure 4.16). The central base pair (2247C-2254G) held slightly longer than the terminus base pair during the simulation with its hydrogen bonding interactions broken up at 80ps (Figure 4.17). The closing base pair (2248G-2253C) appeared to be the most stable base pair and held the longest during high temperature simulation among all three base pairs in the stem. It started to open up after 800ps of simulation (Figure 4.18) indicating that the structure reached the unfolded state.

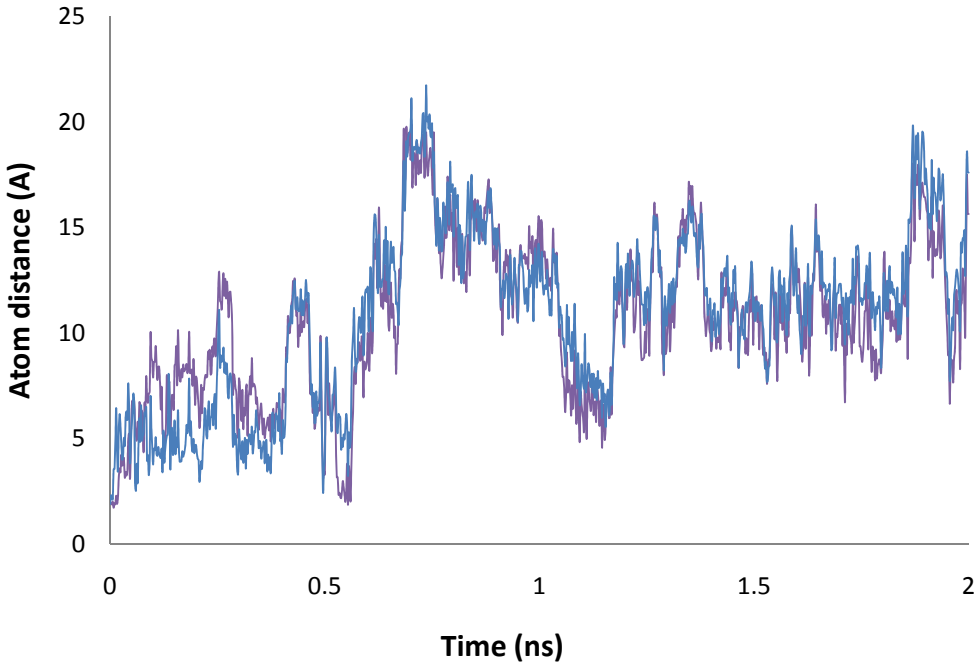



Figure 4.16: Hydrogen bonding interactions between the terminus base pair in the stem during unfolding simulation 2. Distance of O4 on 2246U and H61 on 2255A in purple, distance of H3 on 2246U and N1 on 2255A in blue.

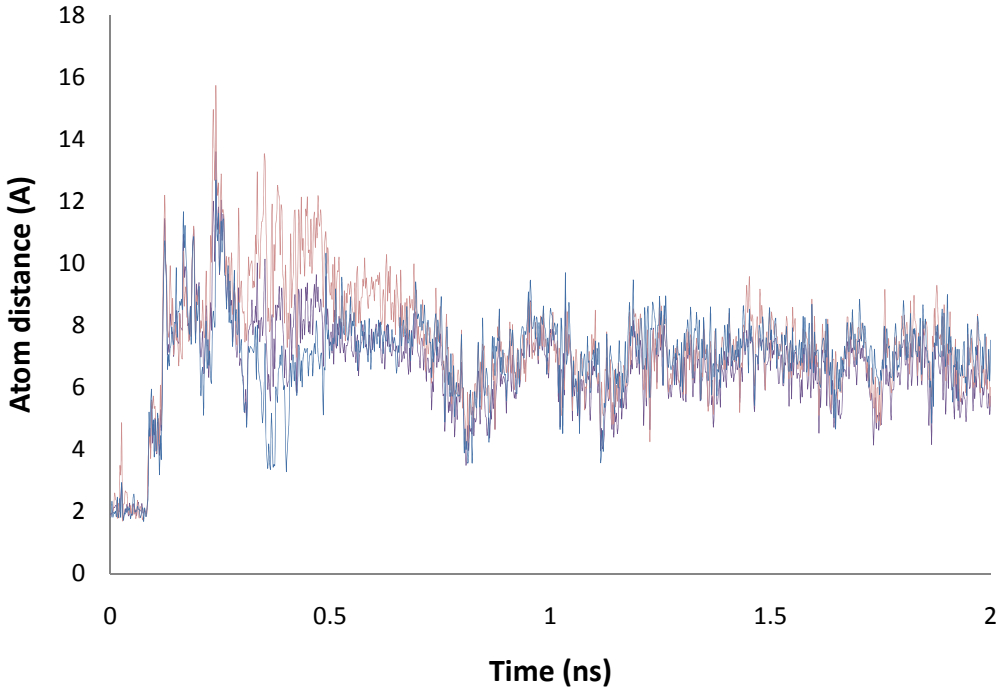



Figure 4.17: Hydrogen bonding interactions between the central base pair in the stem. Distance of N3 on 2247C and H1 on 2254G in purple, distance of H41 on 2247C and O6 on 2254G in pink, and distance of O2 on 2247C and H2 on 2254G in blue.

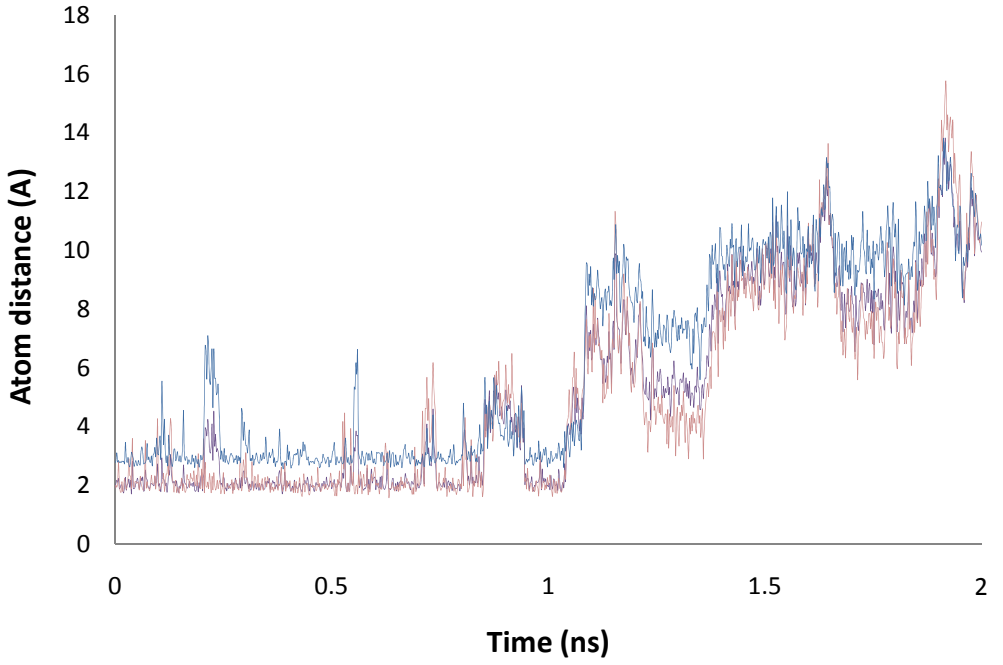



Figure 4.18: Hydrogen bonding interactions between the closing base pair in the stem. Distance of H1 on 2248G and N3 on 2253C in purple, distance of O6 on 2248G and H41 on 2253C in pink, and distance of H21 on 2248G and O2 on 2253C in blue.

The two intra-loop hydrogen bonding interactions, the one between residue j-1 and j+1 and the one between residue j-1 and j+2, were also calculated (Figure 4.19 and Figure 4.20). The intra-loop hydrogen bonding interactions of unfolding simulation 2 were quite different from those of unfolding simulation 1. Recall that during unfolding simulation 1 the intra-loop hydrogen bonding interaction between 2249G (j-1) and 2251G (j+1) was broken soon after the simulation started. In comparison, during unfolding simulation 2 this j-1 – j+1 hydrogen bonding was broken during the first 300ps but reformed from 350ps to 600ps, which corresponded to the intermediate state formation period. This was consistent with the RMSD calculations for residue j-1, j and j+1 in the loop during unfolding simulation 2. Taken together, it quantitatively supported our observation that the unfolding intermediate state of unfolding 2 is structurally similar to a d2 tetraloop.

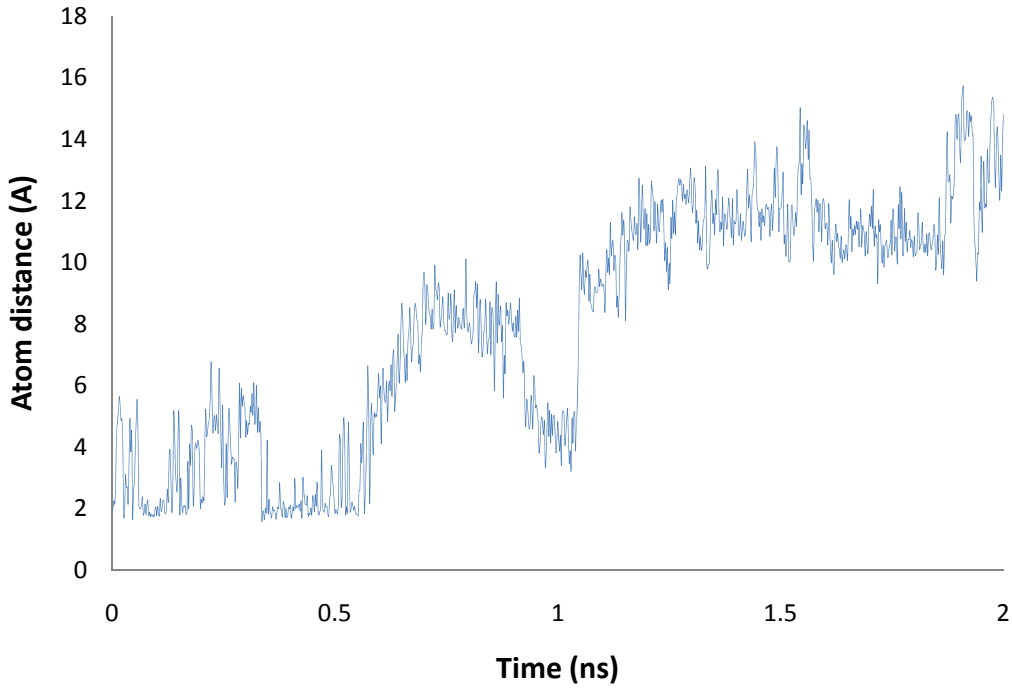



Figure 4.19: Intra-loop hydrogen bonding interactions between 2249G (j-1) and 2251G (j+1) for unfolding simulation 2.

The other intra-loop hydrogen bonding interaction between residue 2249G (j-1) and 2252A (j+2) was conserved for approximately 800ps before it was broken (Figure 4.20). This intra-loop hydrogen bonding interaction between reside j-1 and j+2 was again correlated with the hydrogen bonding interactions between the closing base pair (Figure 4.17) during unfolding simulation 2. This correlation was also observed during unfolding simulation 1. Therefore, the opening up of the closing base pair and the loss of the intra-loop hydrogen bonding interaction between residue j-1 and j+2 indicated that the structure of 2249-standard tetraloop reached its completely unfolded state.

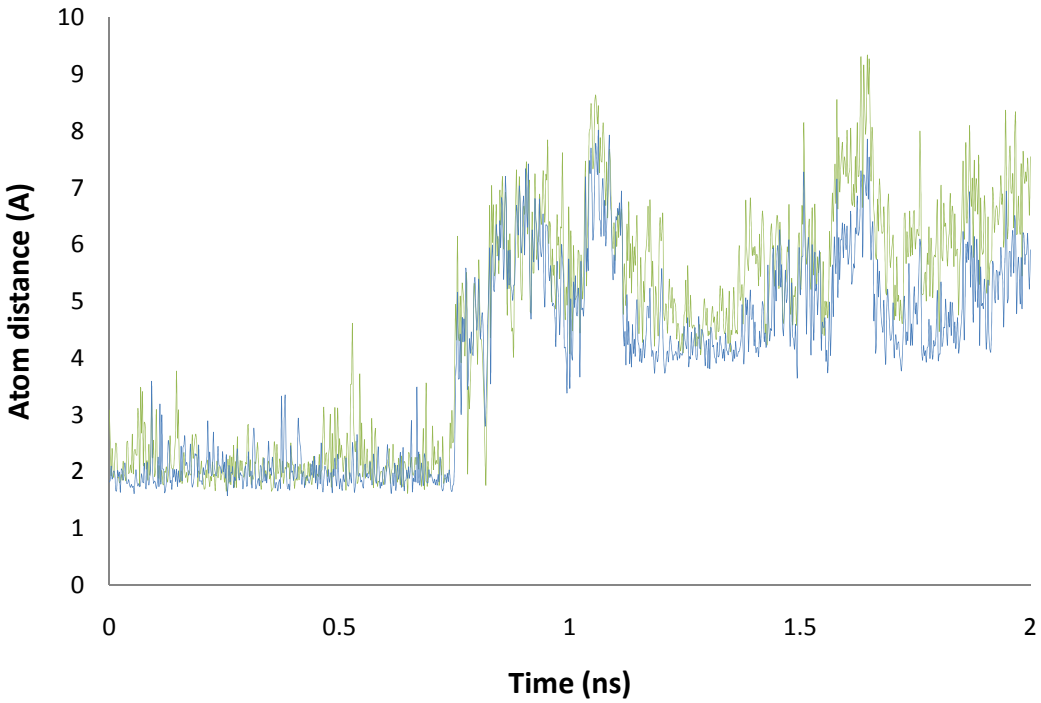



Figure 4.20: Intra-loop hydrogen bonding interactions between 2249G (j-1) and 2252A (j+2), atom distance between H1 of 2249G and O2P of 2252A in green and atom distance between H21 of 2249G and O2P of 2252A in blue.

In summary, hydrogen bonding interactions in the loop and the stem of 2249-standard tetraloop during high temperature simulation 2 showed consistent results with RMSDs analysis. The intermediate state of unfolding pathway 2 was structurally similar to a d2 tetraloop and occurred from 200ps to 500ps. After 500 ps the structure of 2249-standard tetraloop continuously experienced unraveling until the closing base pair opened up completely and the loop deformed completely at 800ps. That was when the structure reached the unfolded state.

*4.3.2.3 Snapshots of Three-States Unfolding Pathway 2: the Folded State, the Intermediate State and the Unfolded State*

High temperature unfolding simulation 2 showed a different folding pathway compared to unfolding simulation 1. Figure 4.21 showed the folded state (left), the intermediate state (center) and the unfolded state (right) during the high temperature simulation 2 of 2249-standard tetraloop.

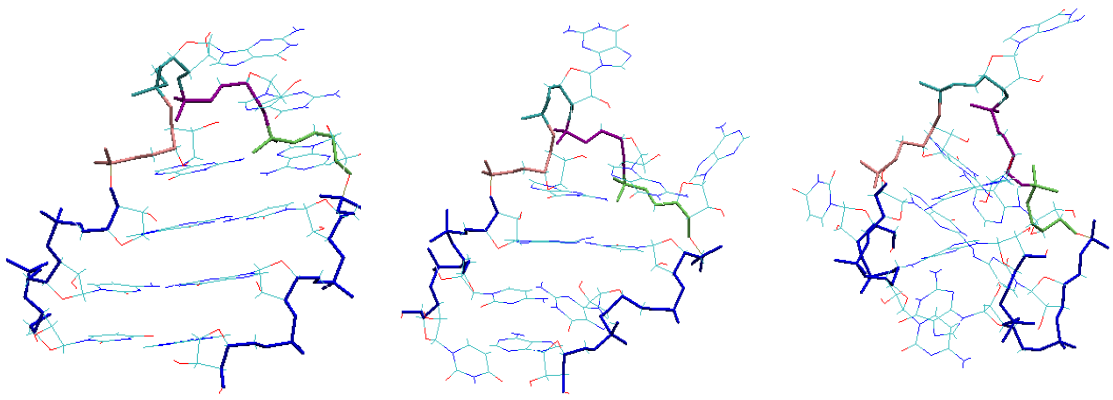


Figure 4.21: Snapshots from unfolding simulation 2. The folded state (starting point) is on the left, the intermediate state (at 500ps) is at the center, the unfolded state (at 1ns) is on the right. The phosphate-ribose backbones of the stem are displayed as blue sticks for clarity. The phosphate-ribose backbones of the loop are colored according to their residue ID, 2249G (j-1) in pink, 2250G (j) in blue, 2251G (j+1) in purple, and 2252A (j+2) in green.

The observed intermediate state (center structure in Figure 4.20) in unfolding pathway 2 had a loop that was structurally very similar to that of a d2 tetraloop, which validated our hypothesis III to some degree. The characterizing intra-loop hydrogen bonding for tetraloop was first broken during the first 300ps of the simulation but reformed at the intermediate state. This hydrogen bonding was absent in the intermediate observed during unfolding simulation 1. The other intra-loop hydrogen bonding interaction between residue j-1 and j+2 was conserved during the first 800ps before it reached the completely unfolded state. This j-1 – j+2 intra-loop hydrogen bonding was correlated with the conserved hydrogen bonding interactions between the closing base pair during the first 800ps.

Despite similarities, the stem of the intermediate state from unfolding simulation 2 was not exactly like that of a d2 tetraloop. The stem of a d2 tetraloop has relatively good base pairing and stacking in place and always has some kind of defects. In the intermediate state of unfolding 2, neither the terminus base pair nor the central base pair was forming hydrogen bonding or stacking. We predicted that during the high temperature unfolding the last residue j+2 in the loop would be shifted down to form closing base pair with j-2, and residue j+3 would be bulged out to become an insertion residue in the stem so that the structure of the intermediate would be like one similar to 314-d2-tetraloop. Apparently the intermediate state was not like what we had predicted.

In unfolding pathway 2, the residue j+2 in the loop of the intermediate state did not shift down to the stem. It was probably because the closing base pair was very stable and could not easily open up. As a result, residue j+2 remained in the loop and was bulged out from its starting point therefore making the rest of the loop become one like a d2 tetraloop. Both RMSDs analyses and hydrogen bonding calculations of the loop supported the observation that the loop of the intermediate state was similar to a d2 tetraloop.

Observation of the simulation trajectory, together with RMSD analysis and

hydrogen bonding interaction calculations, showed that after 800ps the completely unfolded state was reached in unfolding simulation 2. The structure of this unfolded state (right structure in Figure 4.20) was a random coil. Again the structure did not reach its completely unfolded state until the closing base pair started to open up. Therefore, the forming of closing base pair is again considered to be critical for the formation of tetraloops.

## *4.4 Conclusions*

High temperature simulations of 2249-standard tetraloop revealed two unfolding pathways: unfolding pathway 1 and unfolding pathway 2. Both pathways displayed a three-state folding behavior: a native (folded) state, an intermediate state and an unfolded state. Both pathways took similar time to reach the intermediate state and the unfolded state. However the intermediate state of pathway 2 lasted for approximately 300ps, which was shorter the intermediate state of pathway 1 (lasted for approximately 600ps).

In unfolding pathway 2, the structure of the intermediate state resembled that of a d2 tetraloop to some degree, especially the loop. However, we did not observe the residue j+2 in the loop of 2249-standard tetraloop to shift down into the loop forcing residue j+3 to become the insertion residue. We believe it was because the relatively stable closing base pair could not easily open up to allow residue j+2 to shift down from the loop. Instead, residue j+2 was observed to be bulged out from the loop. The structure of the rest of the loop was the same as that of a d2 tetraloop. Therefore, our hypothesis III was validated to some degree.

## *4.5 References*


1. Korostelev A, Trakhanov S, Laurberg M, Noller HF. 2006. Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* 126:1065–77.
2. Bevilacqua, P. C. and Blose J. M. Structures, Kinetics, Thermodynamics, and biological functions of RNA hairpins. *Annu. Rev. Phys. Chem.* 2008 (59): 79 – 103.
3. Tinoco, I. Jr and Bustamante, C. 1999. How RNA folds. *J. Mol. Biol.* 293: 271 – 281.
4. Woese C. 1967. The evolution of the genetic code. In *The Genetic Code: The Molecular Basis for Expression*, pp. 179–95. New York: Harper & Row
5. Crick FH. 1968. The origin of the genetic code. *J. Mol. Biol.* 38:367–79.
6. Orgel LE. 1968. Evolution of the genetic apparatus. *J. Mol. Biol.* 38:381–93.
7. Matzke MA, Birchler JA. 2005. RNAi-mediated pathways in the nucleus. *Nat. Rev. Genet.* 6:24–35.
8. Huttenhofer A, Vogel J. 2006. Experimental approaches to identify noncoding RNAs. *Nucleic Acids Res.* 34:635–46.
9. Nagai K, Oubridge C, Jessen TH, Li J, Evans PR. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* 348:515–20.
10. Munro TP, Kwon S, Schnapp BJ, St Johnston D. 2006. A repeated IMPbinding motif controls *oskar* mRNA translation and anchoring independently of *Drosophila melanogaster* IMP. *J. Cell Biol.* 172:577–88.
11. Gollnick P, Babitzke P, Antson A, Yanofsky C. 2005. Complexity in regulation of tryptophan biosynthesis in *Bacillus subtilis*. *Annu. Rev. Genet.* 39:47–68.
12. Bernstein D, Hook B, Hajarnavis A, Opperman L,Wickens M. 2005. Binding specificity and mRNA targets of a *C. elegans* PUF protein, FBF-1. *RNA* 11:447–58.

13. Babitzke P, Romeo T. 2007. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr. Opin. Microbiol.* 10:156–163.
14. Doudna JA, Cech TR. 2002. The chemical repertoire of natural ribozymes. *Nature* 418:222–28.
15. Fedor MJ,Williamson JR. 2005. The catalytic diversity of RNAs. *Nat. Rev. Mol. Cell Biol.* 6:399–412.
16. Bevilacqua PC, Yajima R. 2006. Nucleobase catalysis in ribozyme mechanism. *Curr. Opin. Chem. Biol.* 10:455–64.
17. Sorin EJ, Engelhardt MA, Herschlag D, Pande VS. 2002. RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop. *J. Mol. Biol.* 317:493–506.
18. Nivon L. G. and Shakhnovich E. I. 2004. All-atom Monte Carlo simulation of GCAA RNA folding. *J. Mol. Biol.* 344: 29 – 45.
19. Brooks, C. L., Simulations of protein folding and unfolding. *Curr. Opin. Struct. Biol*., 1998, Vol. 8: 222-226
20. Pande, V. S. & Rokhsar, D. S., Molecular dynamics simulations of unfolding and refolding of a β-hairpin fragment of protein G. *Proc. Natl Acad. Sci. USA,* 1999, Vol. 96: 9062-9067.
21. Daggett, V. & Levitt, M. Protein unfolding pathways explored through molecular dynamics simulations. *J. Mol. Biol.* 1993, Vol. 232: 600-619
22. Lazaridis, T. & Karplus, M. "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science,* 1997, Vol. 278: 1928-1931
23. Dinner, A. & Karplus, M. Is protein unfolding the reverse of protein folding? A lattice simulation analysis. J. Mol. Biol. 1999, Vol. 292: 403-419
24. Zagrovic, B., Sorin, E. J. & Pande, V. S. β-hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* 2001, Vol. 313: 151-169
25. Jorgensen, W.L., et al., Comparison of simple potential functions for simulating liquid water. Journal of Chemical Physics, 1983. 79: p. 926-935.
26. Cornell, W. D., Cieplak P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M.,

Spellmeyer, D. C., Fox, T., Caldwell, J. W., & Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117 (19): 5179-5197, 1995.

# CHAPTER V

# ROMP POLY(N-UNDECYL-*EXO*-NORBORNENE -5,6-DICARBOXIMIDE)

This chapter presents the preliminary study of the conformational structure of poly(N-undecyl-*exo*-norbornene-5,6-dicarboximide) through bulk simulation to determine if this polymer has a WAXD pattern consistent with a helical conformation.

## *5.1 Introduction*

### 5.1.1 Background

*5.1.1.1 Polymer as support in immobilized catalysts*

Traditional homogeneous and heterogeneous catalysts have been two dominating families of catalysis systems. Homogeneous catalysis occurs when the catalyst and reactants are in the same phase. This type of catalysis is broadly characterized by mild reaction conditions, high yields, and high selectivity. The separation of catalysts from the reaction mixture at the conclusion of reaction can often be difficult on a laboratory scale and nearly impossible on the industrial scale without additional and expensive equipment.

Heterogeneous catalysis occurs when the catalyst and reactants are in different phases. In this type of catalysis, the catalysts are much easier to separate from the products and are more easily regenerated and reused. However, it generally requires

harsher reaction conditions (increased temperature and/or pressure) and increased reaction times while producing lower yields and selectivities than its homogeneous counterpart.

A potential solution to the drawbacks of these two types of catalysis mentioned above is immobilized homogeneous catalysis, commonly called supported or tethered homogeneous catalysis. It is one method that has been used in attempt to develop a catalyst that combines the reactivity and selectivity of homogeneous catalysts with both the separability and reusability of heterogeneous catalysts. Essentially a supported homogeneous catalyst is a homogeneous catalyst which has been attached, or immobilized, on a solid support by a series of atoms that is commonly referred to as either a tether or a linker (Figure 5.1) [1].

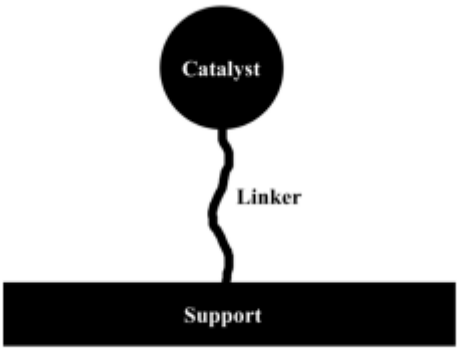



Figure 5.1: A rough sketch of an immobilized homogeneous catalyst [1].

Polymers have previously been used in the application of immobilized homogeneous catalysis and provide many advantages as supports for such systems [1]. In homogeneous catalysis the rate limiting step is reaction rate. Diffusion from the bulk is generally not a concern. With heterogeneous catalysis, diffusion of reactants from the bulk to the catalyst site becomes the rate limiting step. By immobilizing a

homogeneous catalyst, the high reactivity and selectivity of the homogeneous catalyst remains. However, the accessibility of the catalyst to the reactants becomes an issue. Since the catalysts and reactants are still in the same phase, it is not so much diffusion through a phase boundary but diffusion through the support medium that becomes the limiting factor.

In the case of a polymer, a large random coil polymer could block catalyst sites due to the high flexibility of the backbone. The more flexible the backbone, the higher the likelihood of the backbone cutting off a direct route to a catalyst site becomes. The optimal polymer support would be a rigid rod polymer because all of the side groups would be accessible. However, rigid rod polymers are very difficult and expensive to process. Therefore, a useful alternative would be a polymer with a helical conformation if the backbone bonds are flexible, as this adds order to the backbone.

#### *5.1.1.2 Poly(norbornene)*

Poly(norbornene) (PNB) has three possible polymerization pathways, each forming a different polymer, as shown in Figure 5.2 [2]. The first is created via ring-opening metathesis polymerization (ROMP) [3]. The second is created using radical/cationic polymerization [4]. The third is a vinylic polymer created using metal coordinated [5] or metallocene [6] catalysts. ROMP PNB has been used as a polymer support for immobilized homogeneous catalysts, and it is thought that it might share conformational characteristics with its vinylic cousin. The desired characteristic is that of a highly expanded helix which, in turn, could maximize access to the catalytic side groups of the polymer.

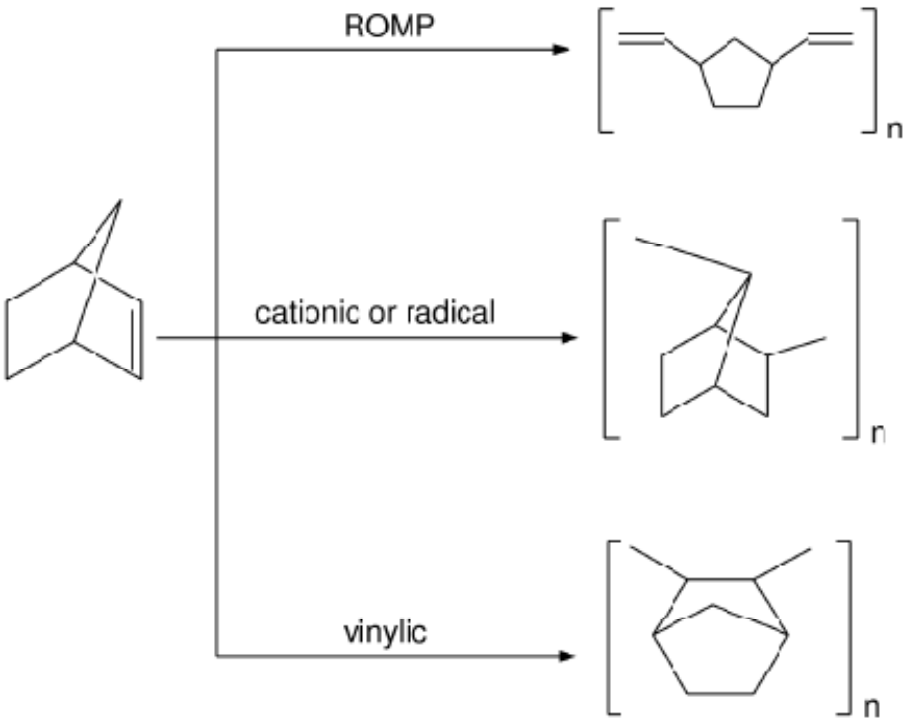



Figure 5.2: The three polymerization pathways of norbornene [2].

Previous work by the Ludovice group on vinylic PNB has shown that for a particular configuration of PNB, the polymer forms a helix-kink conformation. The 2,3-exo,exo-erythro-diisotactic poly(norbornene) configuration was the only one that was found to form this conformation [7]. Because of the helix-kink conformation of this configuration, three hypotheses were formed as to what was required of a helix-kink polymer.

The first hypothesis is that backbone bond types must alternate between rotating and non-rotating bonds. This is consistent in vinylic PNB with the alternating norbornane ring, which is highly constrained therefore cannot rotate, on the backbone of the polymer. The second hypothesis is that the polymer must be cis across the backbone, like exo, exo vinylic PNB. If the polymer were exo, endo, the polymer would not form a helix.

The final hypothesis is that the polymer must have a bulky side group. The best example of this is poly(acetylene). Poly(acetylene) does not have a helical conformation, however, cis-t-butyl poly(acetylene) does. It is believed that the addition of the side group provides the steric hindrance that prevents the torsion angle of the rotatable bond to adopt a more linear conformation. In the case of vinylic PNB, the norbornane ring is playing the role of the bulky side group required to provide the steric hindrance.

One property of this type of kinky helical polymer is its accessibility. If a catalyst is tethered to a polymer support, a random coil polymer has the potential to block the catalyst site to the reactant. A random coil polymer has no underlying structure so it is possible that the catalyst site could be on the interior of the polymer molecule. If a polymer is helical, then there will be an overall order to the polymer backbone. Therefore a catalyst site will be less likely to be inaccessible due to the backbone of the polymer.

ROMP PNB has a very different backbone compared to its vinylic counterpart. Rather than an edge of a norbornane ring as part of the backbone, ROMP PNB contains a cyclopentane ring that is integral to the backbone. However, it is possible that one of the configurations of ROMP PNB would fulfill all of the requirements for a helical polymer. For the polymer itself, there are different possible configurations in the backbone. The double bond can be cis or trans, the backbone can be cis or trans across the cyclopentane ring, and the polymer could have meso or racemic diads (Figure 5.3) [8]. To test the bulky side group hypothesis, a side group can be added to the cyclopentane ring. The side group shown in Figure 5.3 (d) can be found in the literature as a side group which is used to tether catalysts to ROMP PNB [9, 10].

From the aforementioned three requirements for a helical polymer, there is a possibility that ROMP PNB can fulfill all three of them. The first characteristic of ROMP PNB that is important is that the backbone alternates as double bond / single bond / cyclopentane ring. If the cyclopentane ring is rigid enough, this would constitute

the required alternating rotatable / non-rotatable bonds. The cyclopentane ring will always be cis due to the fact that the ROMP mechanism only acts on the double bond of the norbornene ring. As long as the double bond is cis, then the all cis non-rotatable bond requirements will be fulfilled. A side group will be necessary for the third requirement, and racemic diads will ensure that the side group alternates sides of the backbone.

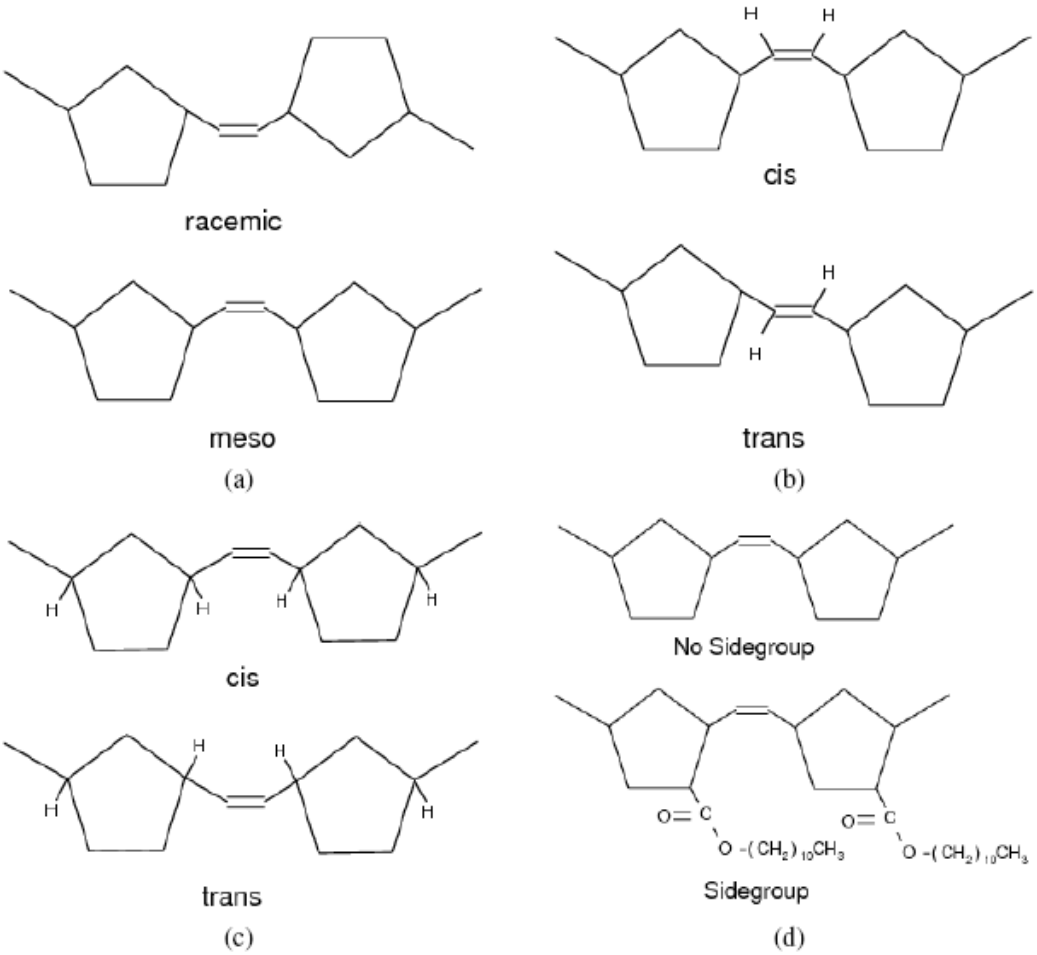


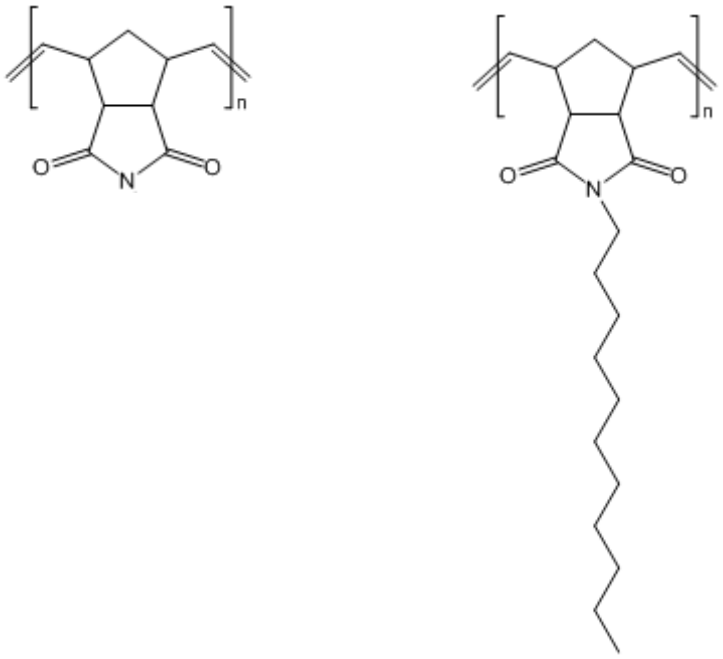

Figure 5.3: The different possible configurations of ROMP PNB. ROMP PNB can have (a) racemic or meso diads, can be (b) cis or trans across the double bond or (c) the cyclopentane ring, and can have (d) an attached sidechain on the cyclopentane ring [8].

Previous modeling of ROMP PNB showed that all of the conformations of ROMP PNB form a random coil. It was determined that the presence of the random coil conformations was due to the fact that the cyclopentane ring on the backbone was not rigid enough to be considered as non-rotatable [8].

#### *5.1.1.3 ROMP poly(N-undecyl-exo-norbornene-5,6-dicarboximide)*

One of the major requirements for a helix-kink polymer is that it must have alternating rotatable / non-rotatable bonds. Previous modeling work showed that ROMP PNB did not form a helical conformation due to the lack of rigidity in the cyclopentane ring in the backbone. However, it is possible to functionalize the cyclopentane ring in such a manner as to remove the flexibility of the backbone cyclopentane ring. Recent experiment studies have shown results that may suggest that the polymerization of a repeat unit with the addition of a second ring fused to the cyclopentane ring could produce a polymer that has a helix-kink conformation [11–15].

ROMP poly(norbornene-5,6-dicarboximide) (PNBDC) (Figure 5.4 *a*) is an example of such polymer with a dicarboximide ring added to the cyclopentane ring to make the ring more rigid so that it can be considered a non-rotatable bond. When an undecyl group was added to the dicarboximide ring the polymer becomes ROMP poly(N-undecyl-exo-norbornene-5,6-dicarboximide) (PNBDC_UD) with a very bulky side group (Figure 5.4 *b*). In ROMP PNBDC or ROMP PNBDC_UD, a second ring is fused to the backbone cyclopentane ring of ROMP PNB to make the cyclopentane ring more rigid and therefore behave as a non-rotatable bond, thus producing a helical conformation. The long side group (undecyl) is also typical of the tether that might be used to attach catalysts.

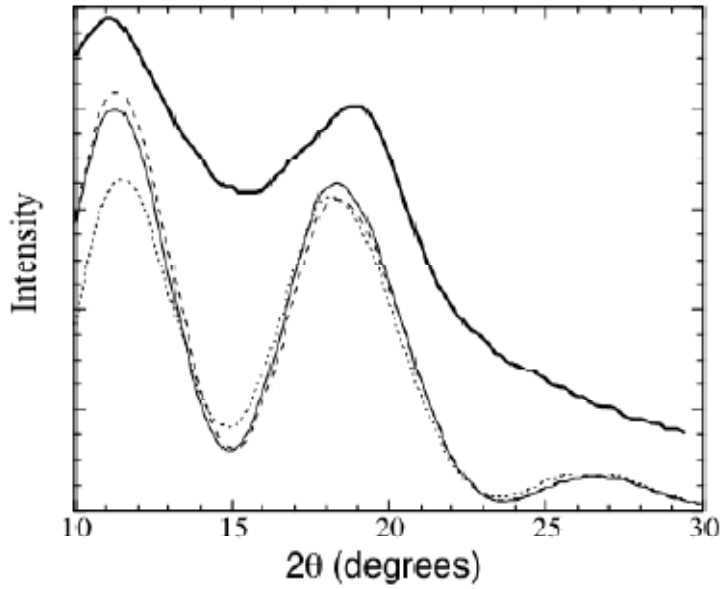



*(a)* *(b)*

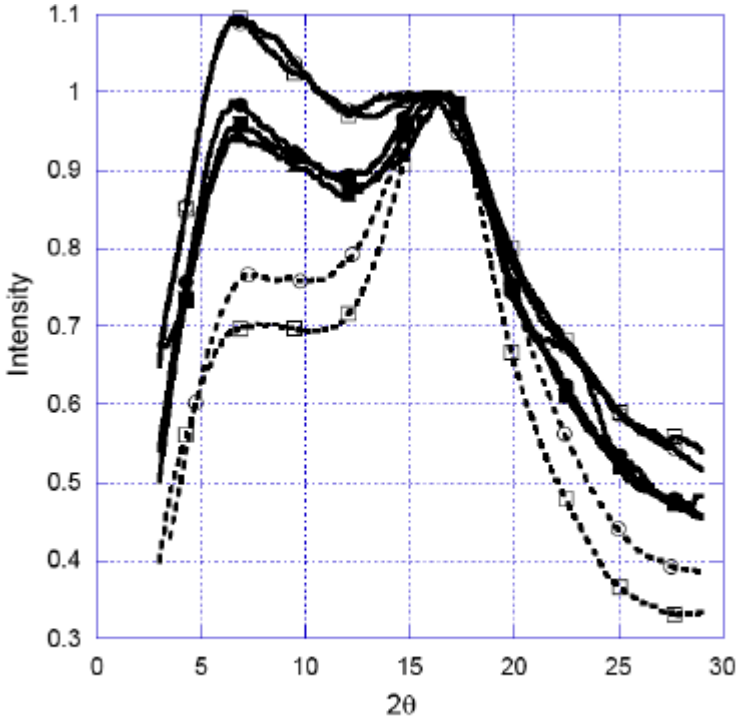
Figure 5.4: Structure of ROMP poly(N-undecyl-exo-norbornene-5,6-dicarboximide).

*5.1.1.4 Wide angle X-ray diffraction (WAXD)*

Wide angle X-ray diffraction (WAXD) is a potential method to characterize the helix-kink type of polymer behavior. The normal diffraction pattern for an amorphous polymer is a broad peak, called an "amorphous halo". For 2,3-exo,exo-erythro-diisotactic PNB, the X-ray diffraction pattern shows two peaks (Figure 5.5).




(a)




(b)

Figure 5.5: WAXD spectra for PNB. (a) The dark solid line represents the experimental WAXD pattern for PNB, the dashed line represents a four chain model of PNB, the solid line represents a six chain model of PNB, and the dotted line represents the WAXD pattern for a poly(ethylene)-PNB 50:50 random copolymer [16]. (b) WAXD spectra for MW from $8.5X10^3$ (top solid line) to $4.74X10^5$ (lowest dotted line) g $mol^{-1}$ [17].

Molecular simulations were found to produce WAXD spectra which matched that of the experimental data (also shown in Figure 5.5(a)). This WAXD pattern can be explained with the helix-kink model that was employed in creating the simulation results. The model shows a split between the intermolecular (low angle) and intramolecular (high angle) interactions from two distinct patterns of order, and further work has shown that the intermolecular order decreases with increasing molecular weight (Figure 5.5 (b)). This is consistent with a helix-kink model where one might expect small, ordered chains that will increase kink content with larger chains. The increased kink content would appear as a decrease in intermolecular order [16, 17].

Molecular modeling has also been used to improve upon the properties of this polymer for its specific application. This was done by determining the effect of different side groups on these same properties of the polymer through modeling. Figure 5.6 shows a comparison of experimental WAXD data with two 100 repeat unit simulations of vinylic PNB with methyl, butyl, and hexyl side groups. There were two peaks for each polymer, but as the side group increased in size, the peak corresponding to the intermolecular portion of the system decreased in intensity. It shows that there will be a change in intermolecular order as the side group increases in size due to the less ordered nature of the larger alkane side groups. This reduction in intermolecular order explains the observed decrease in the glass transition temperature for these polymers.

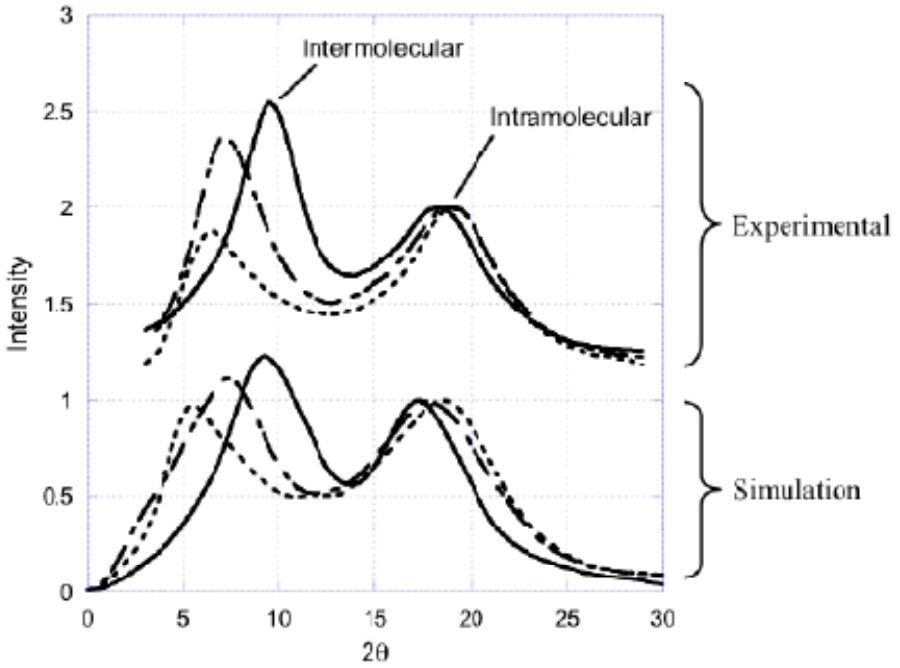



Figure 5.6: Comparison between simulated and experimental WAXD pattern (shifted for comparison) for vinylic PNB with methyl, butyl, and hexyl side groups on the norbornene ring. The simulated and experimental WAXD pattern for the methyl, butyl, and hexyl PNB are represented by the solid line, short-long dashed, and dotted line, respectively [18].

### 5.1.2 Motivation

As discussed earlier one major property of the type of polymer with a helix-kink conformation is its accessibility, which is a major characteristic indicating how good an immobilized catalyst is. The wide angle X-ray scattering experiment conducted by the Weck group generated a WAXD pattern showing that the amorphous halo was split into two distinct peaks, one for intermolecular interactions and the other for intramolecular interactions.

The Ludovice group has identified a certain configuration of ROMP PNBDC_UD has a helical conformation. However bulk simulations have never been carried out on ROMP PNBDC_UD to calculate its WAXD. This work is aimed at

identifying the configuration of ROMP PNBDC_UD that has a helix-kink conformation and performing bulk simulations on the identified configuration to reproduce the experimentally generated WAXD patterns.

### 5.1.3 Objectives

We attempt to reproduce the experimental WAXD pattern for PNBDC_UD using bulk simulations and analyze the resulting order in the simulations. We believe that ROMP PNBDC_UD has some intermediate order.

*Hypothesis IV: ROMP PNBDC_UD forms ordered helices in the bulk similar to non-ROMP PNB.*

To test hypothesis IV, the right combinations of backbone torsion angles that enables the helix-kink conformation of ROMP PNBDC_UD will be determined. A four chain model will be built to perform bulk simulation. WAXD will be calculated and compared with experimental data. Simulations will also be conducted on ROMP PNBDC with no undecyl side group to investigate the role of bulky side group in enabling the polymer to adopt a helical conformation.

## *5.2 Methods*

### 5.2.1 Starting Conformation

Both ROMP PNBDC and ROMP PNBDC_UD with different combinations of backbone torsion angles were used as starting structures in order to identify the appropriate helix-kink conformation and to generate the WAXD pattern that matches the experimental data.

Only two backbone torsion angles, the one before the cyclopentane ring and the one after the cyclopentane ring, were considered rotatable and given different combinations to generate our starting conformation. The first torsion angle was given a value of 120°. The second torsion angle was initially set at a value from 0° to 360° with increment of 20° to provide different combinations of backbone torsion angles. Only isolated polymers with 50 repeating units were simulated as it is the close molecular weight suggested by experiments.

This approach to building a starting conformation gave us a series of possible helical structures, some were broad, some were tight, and some were superhelical. The goal is to run Molecular Dynamics simulations on the polymer with numerous helical starting conformations to see which one of them would stay in helical or at kinky helical conformation. The identified helical or kinky helical conformation would then be used to build bulk models and run bulk simulations so as to generate the WAXD pattern that may match the experimental data.

### 5.2.2 Molecular Dynamics

Prior to each run of molecular dynamics simulation, each starting conformation was minimized with a gradient target of 0.01 kcal $mol^{-1}$ $Å^{-1}$ using the steepest descent algorithm.

NVT molecular dynamics (MD) was used to sample the conformation space at 300K using the Nosé–*Poincaré–Andersen* (NPA) algorithm [19, 20]. The time step was 2 femtoseconds (fs), and the hydrogens were constrained using the LINCS constraint algorithm [21]. The simulations were run for 500 ps. MMFF94 [22, 23, 24] was identified as the most appropriate force field for this polymer and was therefore used as the force field in this study.

### 5.2.3 Wide Angle X-Ray Diffraction

In order to match the WAXD pattern generated experimentally, the identified helical conformation was simulated by placing four polymer chains, each with 50 repeating units, into a periodic crystal cell of P1 symmetry to obtain a density of 0.9 g $cm^{-3}$, which is the estimated density from Weck group. The structure was minimized to the gradient target of 0.01 kcal $mol^{-1}$ $Å^{-1}$.

Due to the fact that ROMP PNBDC_UD has long and bulky side groups that make it very difficult to place four molecule chains in a cell with the right density, a step by step compression protocol was used to reach the right density. The starting crystal cell size had a density of approximately 0.6 g $cm^{-3}$ so that any atom of the four molecule chains would not be placed on top of each other. During each step of compression the size of the cell was reduced by 0.5 Å along x, y, and z dimension respectively and the polymers were simulated by 100ps of NPA MD. The compression was continued until the right density was reached.

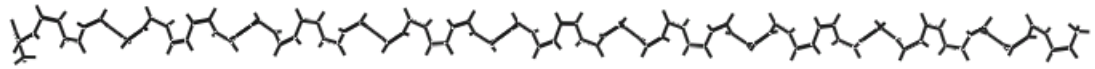
## *5.3 Results and Discussions*

### 5.3.1 Isolated Polymer Simulation

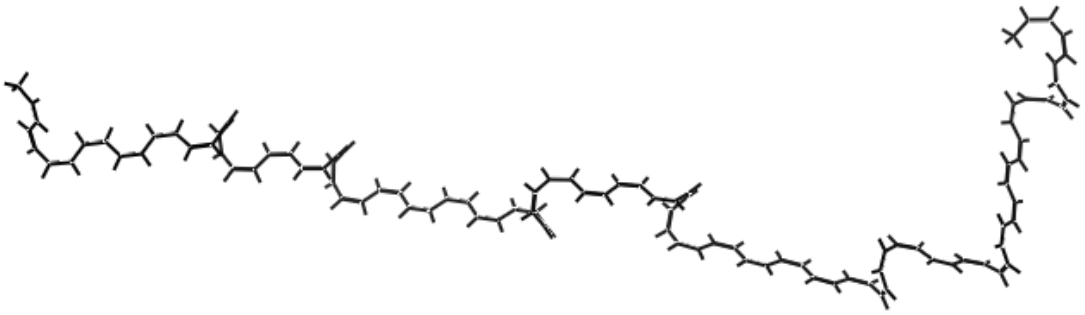
*5.3.1.1 Cis-poly(acetylene) and Cis-t-butyl poly(acetylene)*

It is believed that the addition of the t-butyl side group on poly(acetylene) provides steric hindrance necessary to prevent the torsion angle of the rotatable bond to adopt a more linear conformation. As a starting example to show the role of bulky side groups in making a helix-kink conformation, 500ps MD simulations were conducted at 300K on both *cis*-poly(acetylene) and *cis*-t-butyl poly(acetylene) with a perfect helical starting structure after energy minimization. Each polymer has 50 repeating units.

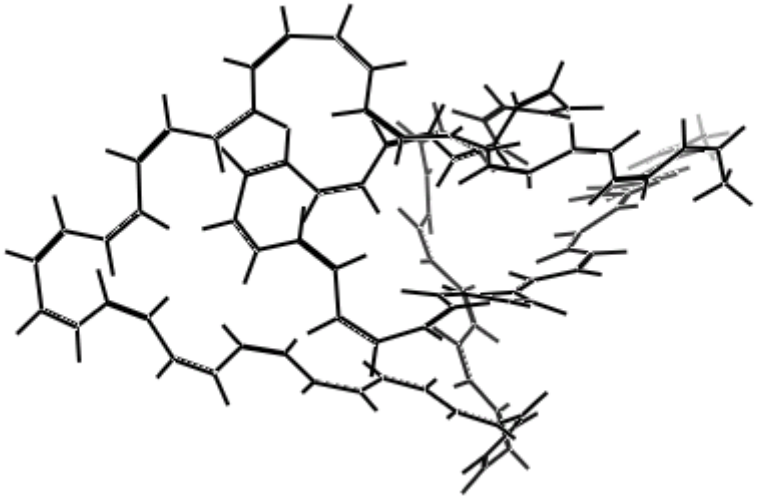
Figure 5.7 and Figure 5.8 showed the structures of both *cis*-poly(acetylene) and *cis*-t-butyl poly(acetylene) with 50 repeating units at the starting point *(a)*, after energy minimization *(b)* and after certain amount of time of MD simulation *(c)*.



*(a)*

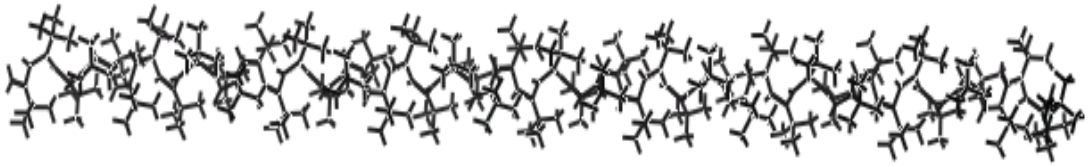


*(b)*

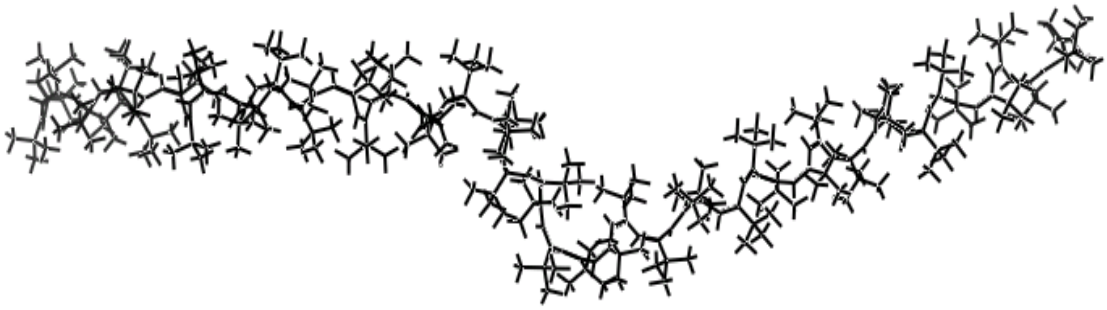


*(c)*

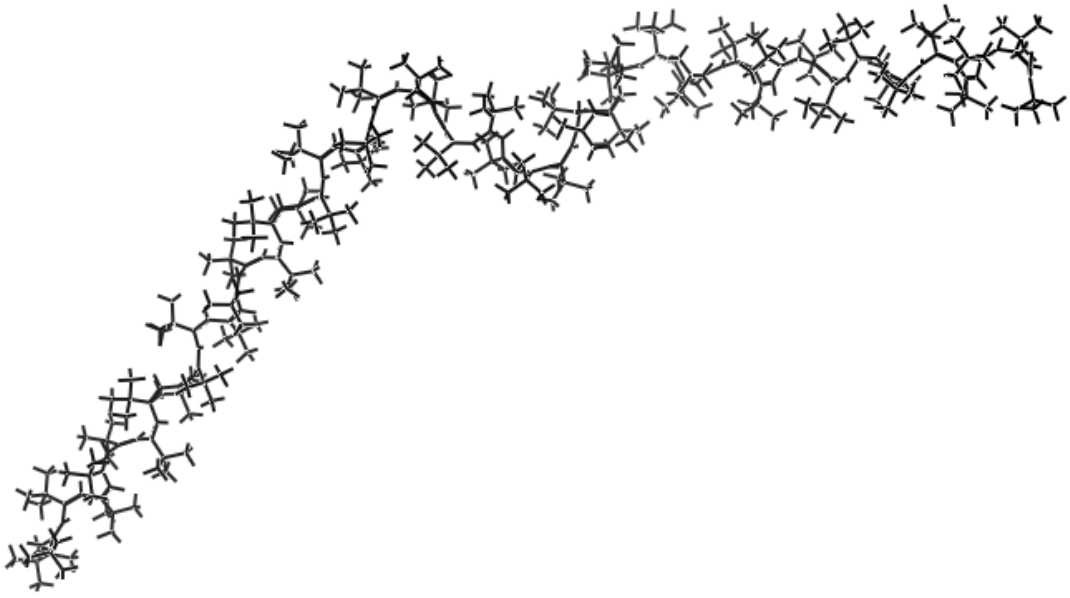
Figure 5.7: Simulation snapshots of *cis*-poly(acetylene). *(a)* helical starting structure, *(b)* structure after minimization, and *(c)* structure after 100ps of MD simulation.



*(a)*



*(b)*



*(c)*

Figure 5.8: Simulation snapshots of *cis*-t-butyl poly(acetylene). *(a)* helical starting structure, *(b)* structure after minimization, and *(c)* structure after 500ps of MD simulation.

Simulation results of *cis*-poly(acetylene) clearly showed that it did not form a helical structure. Rather it formed a random coil after 500ps of MD simulation. In fact *cis*-poly(acetylene) formed a random coil after 100ps of the simulation (Figure 5.7 c). In contrast, *cis*-t-butyl poly(acetylene) with a bulky side group (t-butyl group) added onto the backbone did form a helix-kink conformation even after 1ns of MD simulation (Figure 5.8 c). From these two simulations we could see the role of side groups in making a kinky helical structure.

#### *5.3.1.2 ROMP PNBDC Simulations*

Previous work showed that the cyclopentane ring was not rigid enough to be considered as a non-rotatable bond to make ROMP PNB adopt a helix-kink conformation. This was corrected by adding a fused dicarboximide ring to the cyclopentane ring to increase its rigidity. Previous work has shown that the fused dicarboximide ring is able to effectively limit the motion of the cyclopentane ring and therefore stiffen the backbone cyclopentane ring of ROMP PNB [8]. Therefore, the dicarboximide ring fulfilled the non-rotatable bond requirement for forming a helix. However, it remained unknown whether the dicarboximide group is also large enough to play the role of bulky side groups that also appear to be required for helix formation.

In this study we first investigated the conformation of ROMP PNBDC without the undecyl side group on the dicarboximide ring using MD simulations. Different combinations of backbone torsion angles generated a series of helical structures of ROMP PNBDC, which were minimized energetically. The minimized structures were then simulated at 300K for 500ps. All of the simulations on ROMP PNBDC, no matter what combinations of backbone torsion angles to begin with, generated a random coil conformation at the end of 500ps MD simulation. An example of such random coil conformation of ROMP PNBDC at the end of 500ps MD was exhibited in Figure 5.9.

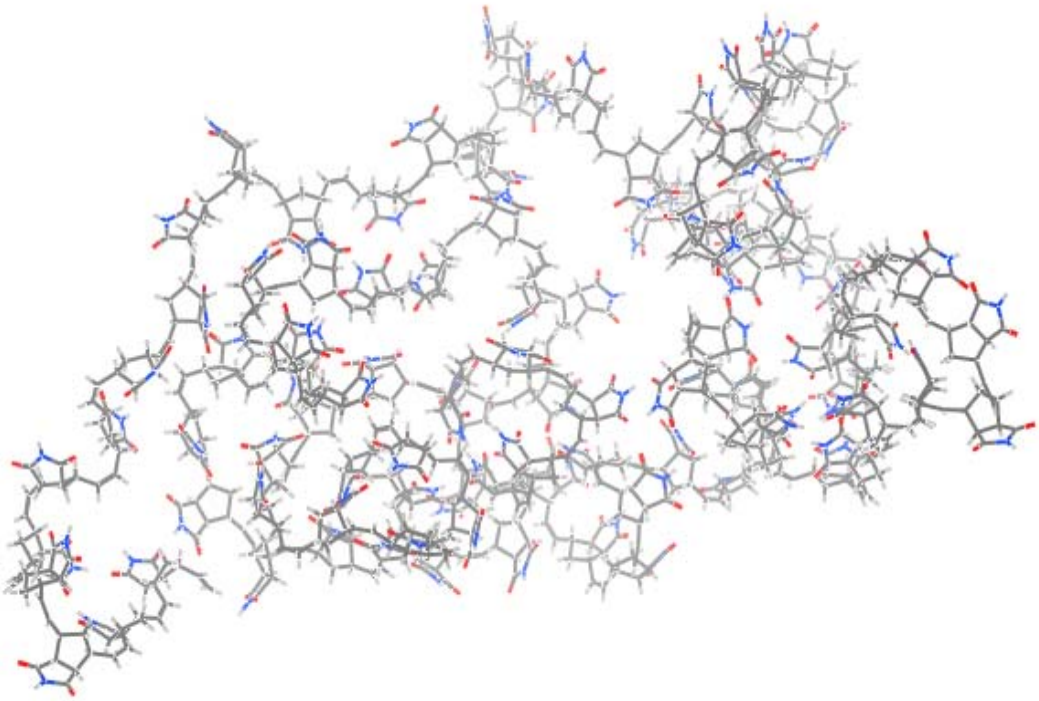


Figure 5.9: End structure of ROMP PNBDC without the undecyl side group on the dicarboximide ring after 500ps MD.

It appeared that although the fused dicarboximide ring could make the cyclopentane ring more rigid, without the undecyl side group it is not bulky enough to provide the steric hindrance that prevents the torsion angle of the rotatable bond to adopt a more linear conformation. Our simulations on ROMP PNBDC without the undecyl side group, no matter what helical conformation it started with, showed that it forms a random coil conformation.

*5.3.1.2 ROMP PNBDC_UD Simulations*

Previous results suggest that the undecyl side group is necessary to provide the steric hindrance for helix formation. Simulations were carried out on ROMP PNBDC with the undecyl side group (ROMP PNBDC_UD) with different combinations of backbone torsion angle as the starting conformation to see if the undecyl side group can fulfill the requirement of the bulky side group for forming a helical structure.

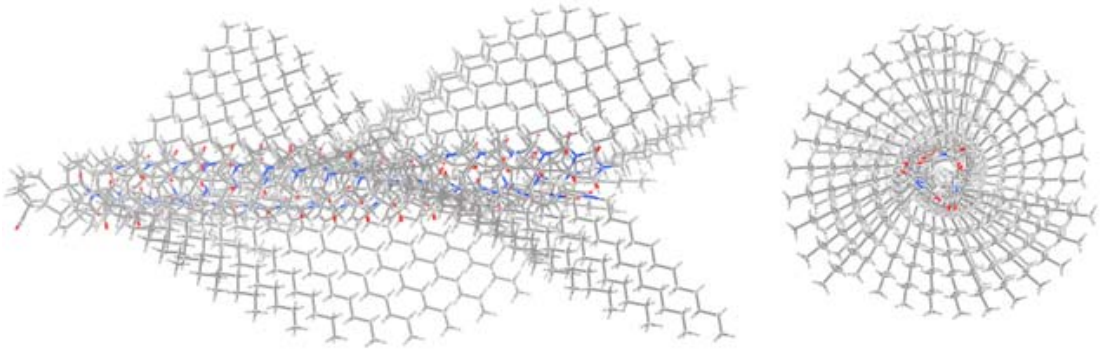
Simulations of ROMP PNBDC_UD showed that two different conformations that adopted helical structures. These two conformations are hereafter labeled as ROMP PNBDC_UD1 and ROMP PNBDC_UD2. Snapshots from simulations of ROMP PNBDC_UD1 and ROMP PNBDC_UD2 were exhibited in Figure 5.10 and Figure 5.11. The rest of ROMP PNBDC_UD with different combinations of backbone torsion angles all form a random coil after 500ps MD simulation.

ROMP PNBDC_UD1 has repeating backbone torsion angles, 120° and 98°, as its starting conformation. The structure of this starting conformation is shown in Figure 5.10 (a). It can be seen that these repeating torsion angles give ROMP PNBDC_UD1 a very tight and highly compact helical structure.

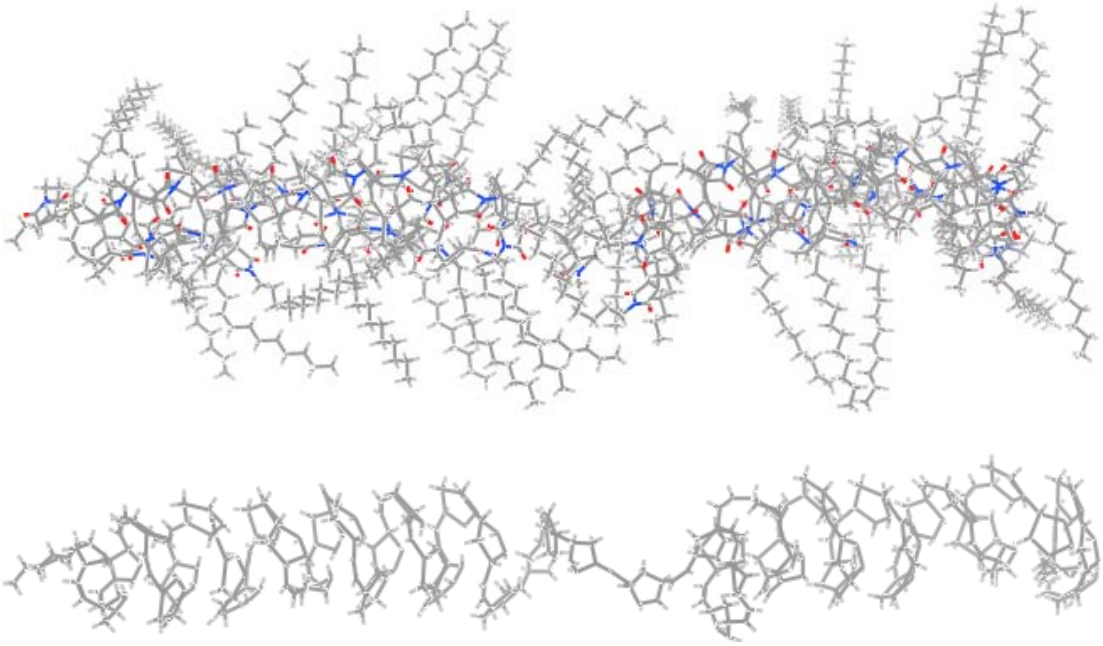
After energy minimization ROMP PNBDC_UD1 remained helical with the compressed starting structure somewhat relaxed, however, a kink similar to the one in the backbone of non-ROMP PNB started to appear in the middle of the chain, which can be seen clearly in the backbone snapshot after minimization (Figure 5.10 (b)). After 500ps of MD simulation, the structure of ROMP PNBDC_UD1 remained helix-kink conformation, which can be seen in the backbone snapshot after the MD simulation (Figure 5.10 (c)).

ROMP PNBDC_UD2 has repeating backbone torsion angles, 120° and 74°, as its starting conformation. The structure of this starting conformation is shown in Figure 5.11 (a). It can be seen that these repeating torsion angles also give ROMP PNBDC_UD2 a very tight helical structure. However, the orientation of the side chain of ROMP PNBDC_UD2 is a little different from that of ROMP PNBDC_UD1.

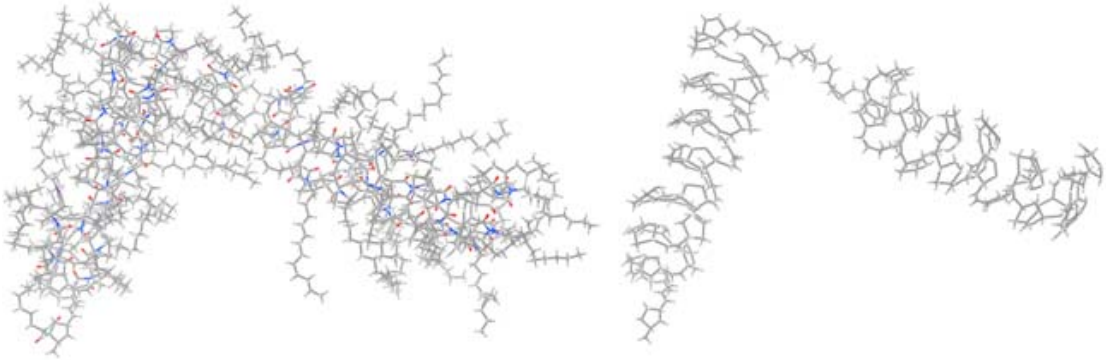
After energy minimization ROMP PNBDC_UD2 helix started to relax from its compressed starting structure (Figure 5.11 (b) backbone snapshot). After 500ps of MD simulation, the structure of ROMP PNBDC_UD2 appeared to be very different from that of ROMP PNBDC_UD1. It remained partially helical, which can be seen in the backbone snapshot (Figure 5.10 (c)), but was much less ordered compared to the structure of ROMP PNBDC_UD1.



*(a)*



*(b)*



*(c)*

Figure 5.10: Simulation snapshots of ROMP PNBDC_UD1. *(a)* helical starting structure, *(b)* structure after minimization (top structure being all atoms and bottom structure being backbone only), and *(c)* structure after 500ps of MD simulation (left structure being all atoms and right structure being backbone only).

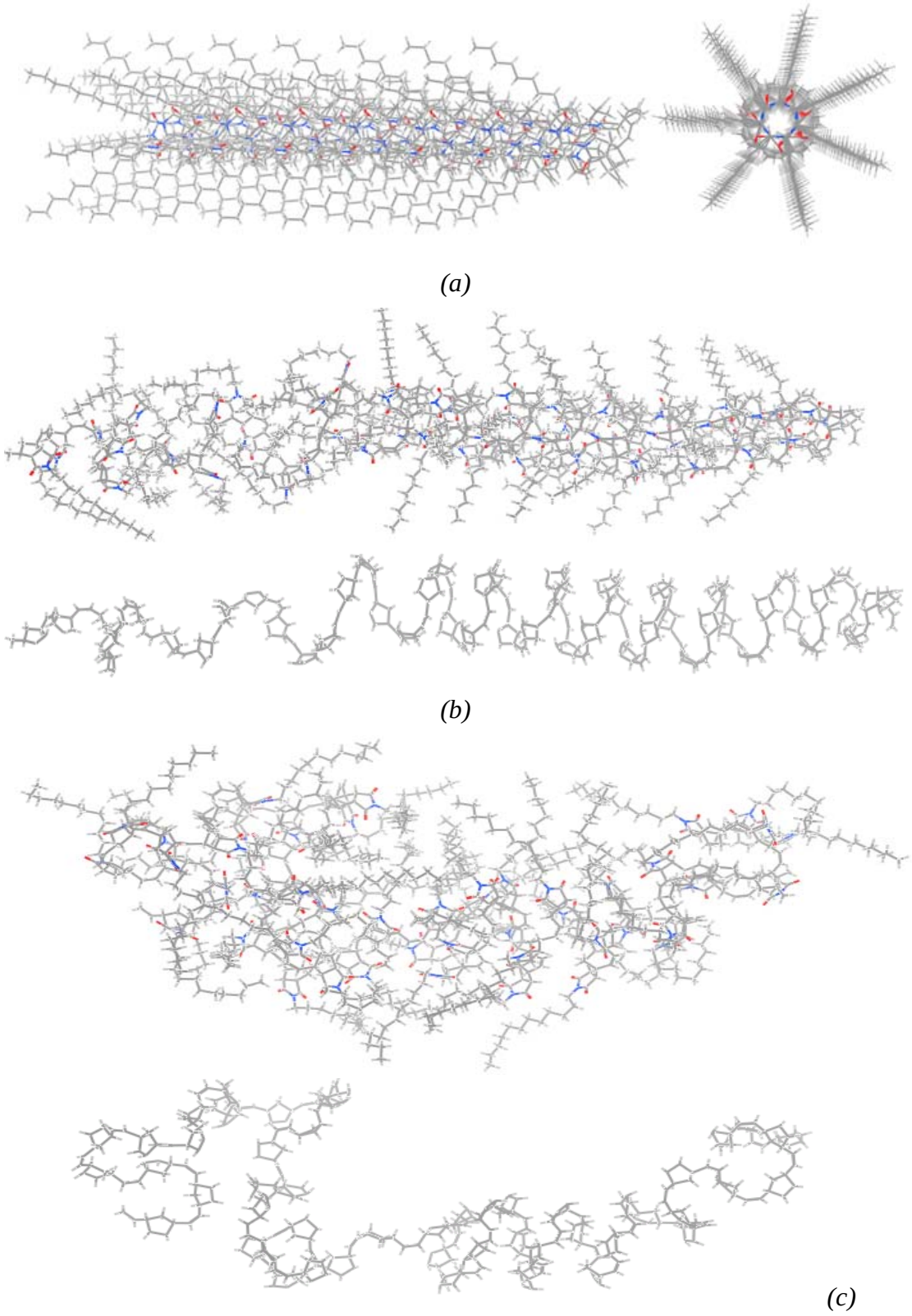


Figure 5.11: Simulation snapshots of ROMP PNBDC_UD2. *(a)* helical starting structure, *(b)* structure after minimization (top structure being all atoms and bottom structure being backbone only), and *(c)* structure after 500ps of MD simulation (left structure being all atoms and right structure being backbone only).

### 5.3.2 Bulk simulation

*5.3.2.1 ROMP PNBDC_UD1 at 300K*

Four molecules of ROMP PNBDC_UD1 with a perfect helical conformation (Figure 5.10 a) were placed in a crystal cell with density of 0.6 g $cm^{-3}$. These four molecules were aligned at the same angle along x, y and z axis respectively. The system reached the right density (0.9 g $cm^{-3}$) following the compression protocol described earlier. After the system was compressed to the right density, the structure was simulated at 300K for 500ps using NPA MD.

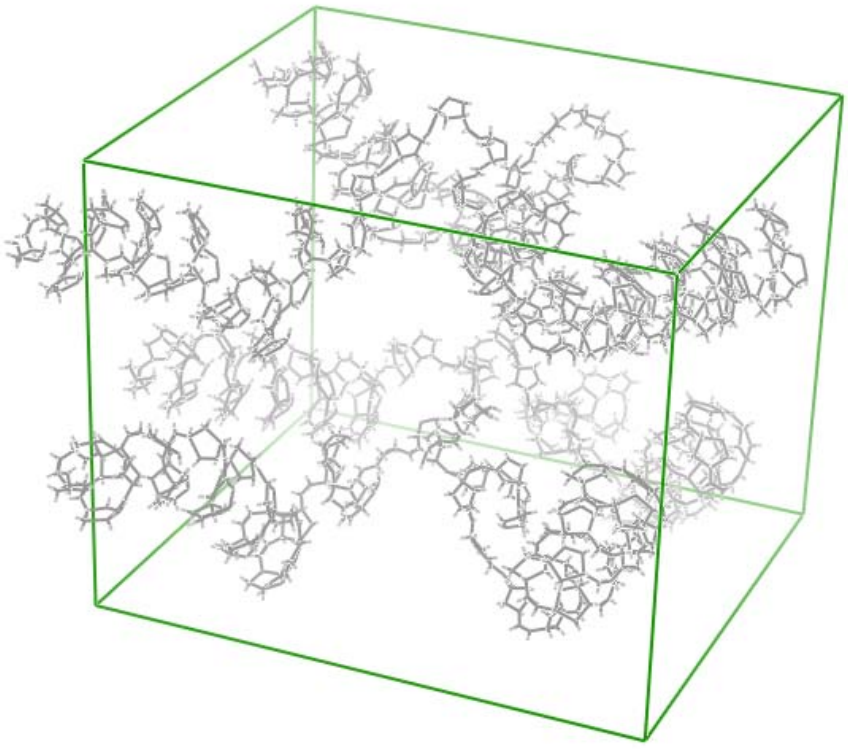


Figure 5.12: Bulk simulation of ROMP PNBDC_UD1. The snapshot of backbone atoms after 500ps MD at 300K, density 0.9 g $cm^{-3}$.

It can be seen from Figure 5.12 that ROMP PNBCD_UD1 has a highly ordered conformation, which is a helix-kink conformation with one kink in the middle of each molecule chain. As mentioned earlier, this kink is similar to the one appeared in the backbone of non-ROMP PNB. The bulk simulation carried out at 300K showed structures very similar to the structure of ROMP PNBCD_UD1 at the end of single molecule MD simulation. The WAXD pattern that this structure generated only showed a single peak at the low angle (calculated in the next section), which represents only the intermolecular interaction.

To make this structure a more amorphous one so that the split peak can be observed, high temperature (900K) bulk simulation of ROMP PNBDC_UD1 was conducted.

*5.3.2.2 ROMP PNBDC_UD1 at 900K*

Previous results showed ROMP PNBDC_UD1 was so ordered that its WAXD pattern only had one single peak at lower angle, which represents the intermolecular interactions. Therefore high temperature simulations were conducted to reduce the structural order. To make sure molecules have enough room to sample conformational space, the starting density was set at 0.25 g $cm^{-3}$ by making the cubic cell size very large. Four molecules of ROMP PNBDC_UD1 placed in the crystal cell were first energetically minimized before compression. The system reached the right density (0.9 g $cm^{-3}$) following the compression protocol described earlier as well. However, what as different was that this time the compression protocol was run at 900K rather than 300K to reduce structural order. After the system was compressed to the right density, the structure was simulated at 900K for 500ps using NPA MD. The end structure of ROMP PNBDC_UD1 after 500ps of bulk simulation was exhibited in Figure 5.13.

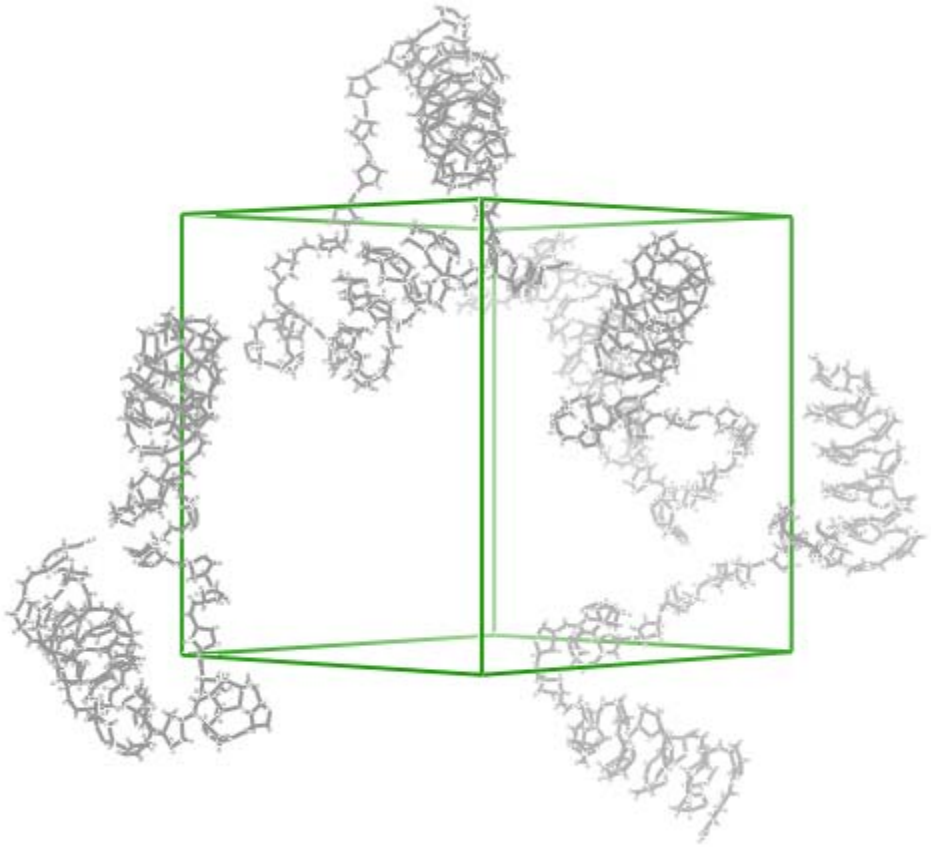


Figure 5.13 High temperature bulk simulation of ROMP PNBDC_UD1, backbone snapshot after 500ps of MD at 900K, density 0.9 g $cm^{-3}$.

Figure 5.13 showed that the structure of ROMP PNBDC_UD 1 after 900K bulk simulation was very similar to that from 300K bulk simulation and remained a highly ordered helix-kink conformation with one kink in the middle of each molecule chain. This conformation again failed to reproduce the WAXD pattern generated experimentally. The WAXD of ROMP PNBDC_UD1 from 900K bulk simulation also only showed a single peak at the lower angle, which suggested that the intermolecular packing of ROMP PNBDC_UD1 conformation is too ordered. Simulations on the different helical conformation, ROMP PNBDC_UD2 were conducted to determine if this starting conformation of ROMP PNBDC_UD can better mimic the experimental results.

*5.3.2.1 ROMP PNBDC_UD 2 at 300K*

A single molecule simulation showed a less ordered structure of ROMP PNBDC_UD2 compared to ROMP PNBDC_UD1. Therefore, ROMP PNBDC_UD2 was used to run bulk simulation and generate WAXD patterns. To run bulk simulation, again four molecules of ROMP PNBDC_UD2, each molecule aligned with the same angle along x, y and z axis respectively, were placed in a crystal cell starting with a density of 0.6 g $cm^{-3}$. The system was minimized first and gradually compressed to the density of 0.9 g $cm^{-3}$ following the aforementioned compression protocol.

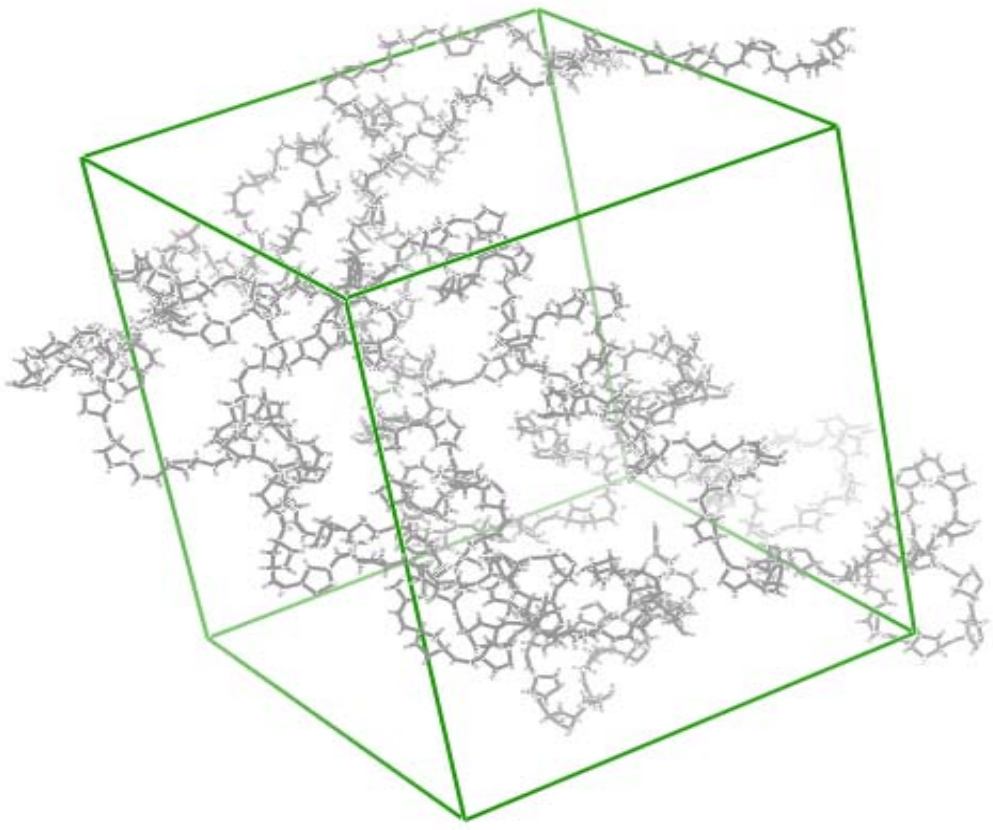


Figure 5.14: Bulk simulation of ROMP PNBDC_UD2. The snapshot of backbone after 500ps MD at 300K, density 0.9 g $cm^{-3}$.

Figure 5.14 shows the backbone snapshot of the structure of ROMP PNBDC_UD2 at the end of 500ps bulk simulation at 300K. It can be seen from Figure 5.14 that the bulk simulation of ROMP PNBDC_UD2 showed a much less ordered conformation when compared to ROMP PNBDC_UD1. The structure of ROMP PNBDC_UD2 had some order in its backbone showing a helical conformation while had the rest of its backbone showing a random coil conformation. In addition, the end structure of ROMP PNBDC_UD2 from the 300K bulk simulation appeared to have less helical regions when compared to the end structure from the single molecule MD simulation (Figure 5.11). The WAXD pattern of ROMP PNBDC_UD2 from the bulk simulation showed consistency with the observation of a less ordered structure and generated two distinct peaks at both lower and higher angle, representing intermolecular and intramolecular interactions respectively (calculated in the next section).

### 5.3.3 Wide angle X-ray diffraction

Figure 5.15 showed the WAXD pattern of ROMP PNBDC_UD with undecyl side group on the dicarboximide ring experimentally generated by the Weck group. The wavelength of the incident radiation used was 1.0903 Å (E = 11 keV) providing a beam flux of greater than $10^{12}$ photons/sec. The two major peaks in the WAXD profile correspond to intersegmental and intrasegmental interactions within the polymer chain, indicating certain order in the conformation of ROMP PNBDC_UD. The peak at the lower angle is due predominantly to the intersegmental interactions, while the peak at the higher angle is due predominantly to the intrasegmental interactions. The positions of lower angle peak and higher angle peak are 2.6° and 13° respectively. All bulk

simulation generated WAXD in this study will be compared to these experimental results.

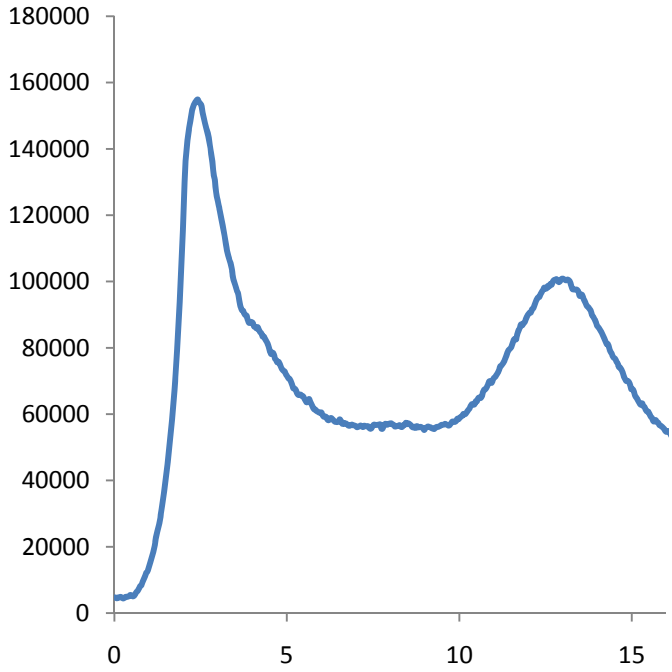



Figure 5.15: WAXD pattern generated by Weck group experimentally, low angle peak at 2.5° , and high angle peak at 12.6° .

*5.3.3.1 WAXD of ROMP PNBDC_UD 1*

As discussed earlier, bulk simulations of ROMP PNBDC_UD 1 showed a highly ordered helix-kink conformation, even when the simulations were carried out at 900K. The calculated WAXD patterns showed consistent results with what were observed during the bulk simulations. Both the 300K simulation and the 900K simulation of ROMP PNBDC_UD1 generated only one single peak at lower angle in WAXD profile indicating only the intermolecular interaction (Figure 5.16 and Figure 5.17). No peak was observed at a higher angle, which corresponds to the intramolecular

interaction.

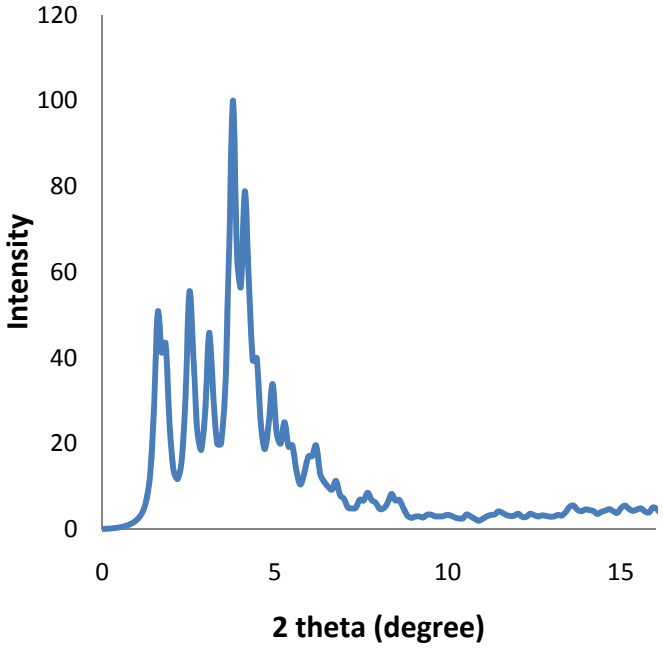



Figure 5.16: WAXD pattern of ROMP PNBDC_UD1 after bulk simulation at 300K. Only a sharp peak at low angle (around 3° – 4°). No peak was observed at the high angle.

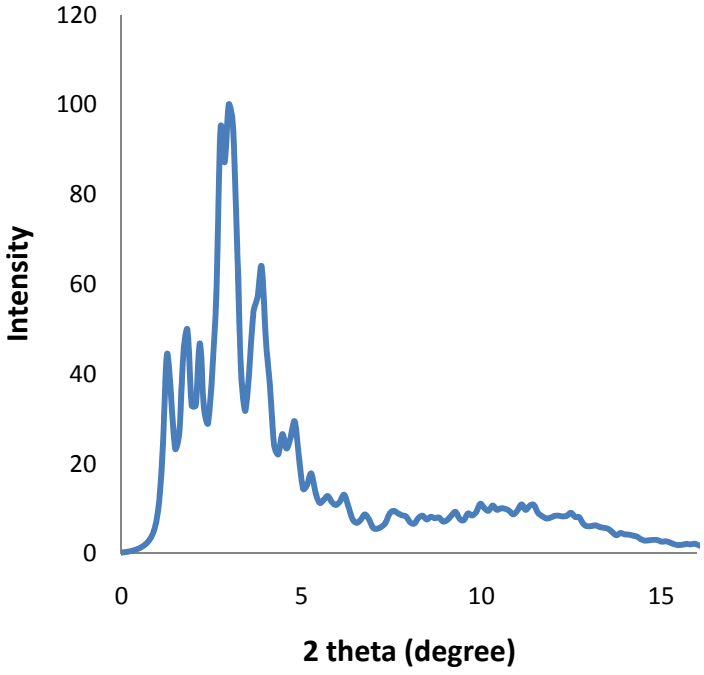



Figure 5.17: WAXD pattern of ROMP PNBDC_UD1 after bulk simulation at 900K. Only a sharp peak at low angle (around 3° – 4°). No peak was observed at the high

angle.

Therefore, the calculated WAXD of simulated ROMP PNBDC_UD1 failed to mimic the experimentally generated WAXD that showed two split peaks at both lower and higher angle (Figure 5.15). The single peak at lower angle in the calculated WAXD profiles suggested that ROMP PNBDC_UD1 had too much order in its conformation.

*5.3.3.2 ROMP PNBDC_UD 2*

The bulk simulation of ROMP PNBDC_UD2 at 300K showed an intermediate ordered conformation where some regions were helical and others were random coil. The calculated WAXD pattern of ROMP PNBDC_UD2 showed two split peaks at both lower and higher angle which were consistent with the experimental results.

Although the calculated WAXD pattern of ROMP PNBDC_UD2 generally matched the experimental data, the intensity had many tiny peaks on the graph and was not very smooth along the x-axis of 2theta. In order to fix the problem and smooth out the calculated WAXD pattern, bulk simulations on ROMP PNBDC_UD2 were repeated 8 times following the exact same simulation protocol. Each repeat generated a WAXD pattern. The WAXD calculated from eight repeats were averaged to generate an averaged WAXD pattern, which was plotted on the same graph with the experimentally generated WAXD pattern for comparison (Figure 5.18).

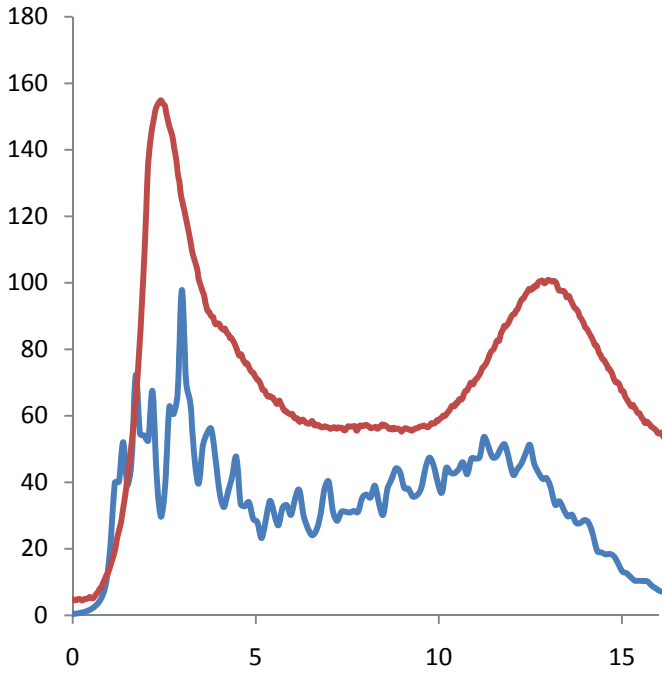



Figure 5.18: Averaged WAXD pattern from 8 repeat of bulk simulations of ROMP PNBDC_UD2 (blue) and WAXD pattern from experiment (red). The low angle peak is at 2-3°, and the high angle peak is at 11.5° for simulated WAXD pattern, while the low angle peak is at 2.5°, and the high angle peak is at 12.6° for experimental WAXD pattern.

Figure 5.18 showed that the simulated WAXD pattern of ROMP PNBDC_UD2 generally matched the experimental WAXD pattern. The simulated WAXD pattern had several small splits of peaks at lower angle, which could be some artifacts of the small system. The simulated WAXD pattern had its higher angle peak at 11.5° while the experimental WAXD pattern had its higher angle peak at 12.6°. This could be because the density used in the simulation was an estimate and may not be the exact same density of ROMP PNBDC_UD that was used to generate the WAXD pattern experimentally.

In summary, both the structure from bulk simulation at 300K and the calculated WAXD pattern of ROMP PNBDC_UD2 showed an intermediate conformation which appeared to be consistent with the experimental results. The

similarity between the experimental results and bulk simulations of ROMP PNBDC_UD2 suggest that a helical conformation may occur in ROMP PNB when adding a dicarboximide ring with a long bulky side group to the cyclopentane ring.

## *5.4 Conclusions*

Simulations of ROMP PNBDC showed a random coil conformation and suggested that only the fused dicarboximide ring without the bulky side group (undecyl) was not enough to make the structure to adopt a conformation with order. Simulations of ROMP PNBDC_UD showed two versions of helix-kink conformation with different backbone torsion angles: ROMP PNBDC_UD1 and ROMP PNBDC_UD2. Bulk simulations of ROMP PNBDC_UD1 showed a structure with too much order therefore failed to generate the WAXD pattern that matches the experimental data. Bulk simulations of ROMP PNBDC_UD2 showed a conformation with an intermediate order and were able to generate the WAXD that is close to the experimental data.

In summary, this work showed preliminary results on the important role of bulky side groups for helix formation. While the fused dicarboximide ring makes the cyclopentyl group non-rotatable, the undecyl side group is required to provide the steric hindrance that is capable of making ROMP PNB rigid and adopt a conformation with more order that match the observed WAXD in ROMP PNBDC_UD.

## *5.5 References*


1. Clark, J. H. and Macquarrie, D. J., Kirk-Othmer Encyclopedia of Chemical Technology Online, vol. 5, ch. Supported Catalysts, pp. 322–344. Wiley-Interscience, 2002.
2. Karafilidis, C., Hermann, H., Rufinska, A., Gabor, B., Mynott, R. J., Breitenbruch, G.,Weidenthaler, C., Rust, J., Joppek,W., Brookhart, M. S., Thiel,W., and Fink, G., "Metallocene-catalyzed c7-linkage in the hydrooligomerization of norbornene by sigma-bond metathesis: Insight into the microstructure of polynorbornene," *Angewandte Chemie-International Edition*, vol. 43, no. 18, pp. 2444–2446, 2004.
3. Ivin, K. J. and Mol, J. C., *Olefin Metathesis and Metathesis Polymerization.* Academic Press, San Diego, 1997. p. 407.
4. Gaylord, N. G., Deshpande, A. B., Mandal, B. M., and Martan, M., "2,3- and 2,7-bicyclo[2.2.1]hept-2-ene. preparation and structures of polynorbornene," *Journal of Macromolecular Science and Chemistry*, vol. 11, pp. 1053–1070, 1977.
5. Goodall, B., Benedikt, G., McIntosh, L., Banres, D., and Rhodes, L., "Novel heat resistant cyclic olefin polymers made using single component nickel and palladium catalysts," *Polymeric Materials Science and Engineering*, vol. 76, pp. 56–57, 1997.
6. Kaminsky, W. and Noll, A., "Copolymerization of norbornene and ethene with homogeneous zirconocenes methylaluminoxane catalysts," *Polymer Bulletin*, vol. 31,no. 2, pp. 175–182, 1993.
7. Chung, W. J., *Molecular modeling of structure-property relationship for palladium catalyzed poly(norbornene) and its derivatives.* PhD thesis, Georgia Institute of Technology, 2003.
8. Swann, A. T., *Characterization of polymer-supported homogeneous catalysts by molecular modeling.* PhD thesis, Georgia Institute of Technology, 2008.
9. Pollino, J. M., Stubbs, L. P., and Weck, M., "Living romp of exo-norbornene esters possessing pdii scs pincer complexes or diaminopyridines," *Macromolecules*, vol. 36, pp. 2230–2234, 2003.
10. Carlise, J. R. andWeck, M., "Side-chain functionalized polymers containing

bipyridine coordination sites: polymerization and metal-coordination studies," *Journal of Polymer Science Part A: Polymer Chemistry*, vol. 42, pp. 2973–2984, 2004.

11. Sattigeri, J. A., Shiau, C. W., Hsu, C. C., Yeh, F. F., Liou, S., Jin, B. Y., and Luh, T. Y., "Remarkable enhancement of second-order nonlinear optical properties of polynorbornenes having pendant chromophores. use of hyper-rayleigh scattering to estimate the tacticity of rigid rod polymers," *Journal of the American ChemicalSociety*, vol. 121, pp. 1607–1608, Feb. 1999.
12. Lin, W. Y., Murugesh, M. G., Sudhakar, S., Yang, H. C., Tai, H. C., Chang, C. S., Liu, Y. H., Wang, Y., Chen, I. W. P., Chen, C. H., and Luh, T. Y., "On the rigidity of polynorbornenes with dipolar pendant groups," *Chemistry-A European Journal,* vol. 12, pp. 324–330, 2006.
13. Diaz, K., Vargas, J., Del Castillo, L. F., Tlenkopatchev, M. A., and Aguilar-Vega, M., "Polynorbornene dicarboximides with cyclic pendant groups: synthesis and gas transport properties," *Macromolecular Chemistry and Physics*, vol. 206, pp. 2316–2322, 2005.
14. Contreras, A. P., Cerda, A. M., and Tlenkopatchev, M. A., "Synthesis of hightg polymers by ring-opening metathesis polymerization of n-cycloalkylnorbornene dicarboximide," Macromolecular Chemistry and Physics, vol. 203, pp. 1811–1818, 2002.
15. Hino, T., Inoue, N., and Endo, T., "Ring-opening metathesis copolymerization behaviors of cyclooctene and norbornene bearing a five- or six-membered ring cyclic carbonate," *Journal of Polymer Science Part A-Polymer Chemistry,* vol. 43, p. 6599, Dec. 2005.
16. Ahmed, S., Ludovice, P. J., and Kohl, P., "Microstructure of 2,3 erythro di-isotactic polynorbornene," *Computational and Theoretical Polymer Science,* vol. 10, pp. 221–233, 2000.
17. Callander, D. B., *Molecular modeling of polymer free volume distribution. PhD thesis*, Georgia Institute of Technology, 2005.

18. Wilks, B. R., Chung, W. J., Ludovice, P. J., Rezac, M. E., Meakin, P., and Hill, A. J., "Structural and free-volume analysis for alkyl-substituted palladium-catalyzed poly(norbornene): A combined experimental and monte carlo investigation," *Journal of Polymer Science Part B-Polymer Physics,* vol. 44, pp. 215–233, Jan. 2006.
19. Bond, Stephen, D., Leimkuhler, Benedict, J., and Laird, Brian, B.. The nose-poincare method for constant temperature molecular dynamics. *J.Comput. Phys.*, vol. 151, no. 1, pp. 114{134, 1999. 316646.
20. Sturgeon, J. B. and Laird, B. B. Symplectic algorithm for constantpressure molecular dynamics using a nose-poincare thermostat. *Journal of Chemical Physics*, vol. 112, no. 8, pp. 3474{3482, 2000.
21. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M., "Lincs: A linear constraint solver for molecular simulations," Journal of Computational Chemistry, vol. 18, pp. 1463–1472, Sept. 1997.
22. Thomas A. Halgren. Merck molecular force field. I. Basis, form, scope, parameterization and performance of MMFF94. *J. Comput. Chem.* 17: 490-519, 1996.
23. Thomas A. Halgren. Merck molecular force field. II. MMFF94 van der waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* 17: 520-552, 1996.
24. Thomas A. Halgren. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem*. 17: 553-586, 1996.

# CHAPTER VI

# CONCLUSIONS AND RECOMMENDATIONS

## *6.1 Conclusions*

Molecular modeling has been used to investigate the structures, dynamics and thermodynamics of biological systems and polymeric systems. This work has applied Molecular Dynamics (MD) simulations to study the secondary structure of both a biological molecule, Ribonucleic Acid (RNA), and a synthetic polymer, ROMP poly(N-undecyl-exo-norbornene-5,6-dicarboximide) (ROMP PNBDC_UD). The work on RNA tetraloops enriches the understanding of the thermodynamics and folding mechanism of this very important RNA motif. The work on ROMP PNBDC_UD contributes to the general design of this polymer as homogeneous catalysts support with improved accessibility. From this work the following conclusions may be drawn:

1. A d2 tetraloop with no defects in the stem, or the perfect d2 tetraloop is not a stable structure.

   This was confirmed by comparing the MD simulations of a standard tetraloop (805-standard tetraloop), a d2 tetraloop (314-d2-tetraloop) and a perfect d2 tetraloop (cutd2-tetraloop). Simulation observations, RMSD analyses, hydrogen bonding calculations, and torsion angle showed that an unstable perfect d2 tetraloop relative to the standard tetraloop and d2 tetraloop. This is consistent with the fact that the perfect d2 tetraloop has never been observed in the database and suggests that stem defects are important for stabilizing the structure of d2 tetraloop.

2. The stem defect appears to be a stabilizing factor for d2 tetraloops.

   The removal of the insertion residue from 314-d2-tetraloop made an unstable structure, cutd2-tetraloop. The simulation of 314-d2-tetraloop showed the insertion formed two hydrogen bonding interactions: one with the loop and one with the stem. Therefore the insertion residue is contributing to the stability of 314-d2-tetraloop by decreasing the enthalpy of the system. Further investigation of the role of the insertion residue by eliminating these two hydrogen bonding interactions showed that the structure remained relatively stable especially in the stem. This suggests that the insertion residue also reduces conformational strain and therefore contributes to the stability of 314-d2-tetraloop.

   Simulations suggested the stem defects have to be structurally flexible in order to provide stabilizing hydrogen bonding interactions with the rest of the structure. This is consistent with the observations of stem defects of all the d2 tetraloops in the database. All the stem defects are flexible structure, i.e. none of them is a rigid stem. Therefore, while our simulations were carried out on one specific d2 tetraloop, 314-d2-tetraloop, this insertion – stabilization is consistent with other d2 tetraloops.

3. The folding of tetraloops starts with the formation of the loop and may have the d2 tetraloop as an intermediate state.

   High temperature simulation of 2249-standard tetraloop revealed two unfolding pathways, both of which showed that the folding of tetraloops started with the formation of loop. The closing base pair formed first, and then the loop and the stem started to form the critical interactions such as base pairing and stacking that make a tetraloop and its associated stem. Both unfolding pathways showed an intermediate state. In unfolding pathway 2, the structure of the intermediate state resembled to a large extent that of a d2 tetraloop, suggesting that the d2 tetraloop

may be a kinetically trapped intermediate during the folding of the standard tetraloop. However the intermediate was not exactly the same as a d2 tetraloop with an insertion such as 314-d2-tetraloop as we predicted. It was simply because the extreme stable closing base pair that could not easily open up. As a result the fourth residue (j+2) in the loop of 2249-standard tetraloop could not shift down to the stem to become j+3 and push the residue j+3 out to become the insertion residue.

4. The undecyl bulky side group plays an important role in making ROMP PNBDC adopt a helix-kink conformation.

   Simulations of ROMP PNBDC showed a random coil conformation and suggested that only the fused dicarboximide ring without the bulky side group (undecyl) was not enough to make the structure to adopt a conformation with order. Simulations of ROMP PNBDC_UD showed two versions of helix-kink conformation with different backbone torsion angles: ROMP PNBDC_UD1 and ROMP PNBDC_UD2. Bulk simulations of ROMP PNBDC_UD1 showed a structure with too much order therefore failed to generate the WAXD pattern that matches the experimental data. Bulk simulations of ROMP PNBDC_UD2 showed a conformation with an intermediate order and were able to generate the WAXD that is very close to the experimental data.

## *6.2 Recommendations and Future work*

1. 314-d2-tetraloop, a sample structure of a d2 tetraloop with only one insertion residue in the stem was used to test the effect of stem defects. This is for the purpose of making the system computationally less expensive. For a better understanding of the effect of stem defects, it is recommended that future work investigate d2

tetraloops with more than one insertion residues or with other types of defects in the stem.

2. Four runs of high temperature unfolding simulations of 2249-standard tetraloop were conducted in this work, which generated two different folding pathways. More simulations could be done on the unfolding simulations of standard tetraloops to confirm the structure of the intermediate state of folding.
3. 2249-standard tetraloop were used as a sample structure of standard tetraloops to study its folding. More standard tetraloop sample structures could be tested under the same simulation method.
4. High temperature unfolding simulations were used in this work in order to reveal folding pathway of a standard tetraloop. However, the approach of high temperature unfolding simulation may produce artifacts in the simulation. It is recommended that other types of biasing methods could be used to explore the folding of tetraloops, such as Steered MD [1], Maxwell's demon MD [2], and replica-exchange MD. Folding pathway generated from different methods can be compared with each other to minimize simulation artifacts.
5. The conformational search of ROMP PNBDC_UD backbone torsion angle was manually conducted by fixing one backbone torsion angle fixed at 120° and setting the other at a value from 0° to 360° with increment of 20°. Further investigation on the first backbone torsion angle is recommended to explore more possible conformations.
6. The calculated WAXD pattern of ROMP PNBDC_UD2 was close to the experimental data. However, it was not smooth enough and had a lot of small peaks. This could be further average out by different approaches.

## *6.3References*


1. Israelewitz, B., S. Izrailev, and K. Schulten, Binding pathway of retinal to bacterioopsin: a prediction by molecular dynamics simulations. *Biophys J*, 1997. 73(6): 2972-9.
2. Harvey, S.C. and H.A. Gabb, Conformational transitions using molecular dynamics with minimum biasing. *Biopolymers*, 1993. 33(8): p. 1167-72.
3. *Sugita, Yuji; Yuko Okamoto (1999). "Replica-exchange molecular dynamics method for protein folding". Chem Phys Letters **314**: 141–151.*